AD 696 152

A COMPUTER SIMULATION OF AUDIENCE EXPOSURE IN A MASS MEDIA SYSTEM: THE UNITED NATIONS INFORMATION CAMPAIGN IN CINCINNATI, 1947-1948

John Francis Kramer

Massachusetts Institute of Technology
Cambridge, Massachusetts

June 1969

C/69-11

A COMPUTER SIMULATION OF AUDIENCE EXPOSURE IN A MASS MEDIA SYSTEM: THE UNITED NATIONS INFORMATION CAMPAIGN IN CINCINNATI, 1947-1948

John Francis Kramer

Research Program on Problems of International
Communication and Security

Center for International Studies
Massachusetts Institute of Technology
Cambridge, Massachusetts

Ju[illegible] 1969

PREFACE

Anyone who has worked in education or mass communication knows how hard it is to change men's minds or arouse their interest. A classic study of the problem was done in Cincinnati on a United Nations Association campaign; Dr. Kramer examines it in this monograph. But there are many other such studies. Voting by Bernard Berelson, Paul Lazarsfeld, and William McPhee established that few people are converted by a Presidential campaign.* Studies of psychological warfare, on which we shall comment more fully below, show too that defectors are won only from among already alienated troops. The notion that "Hidden Persuaders" can by a few devious tricks manipulate large publics by mass media has long since been exposed as a myth, the widespread survival of which illustrates once more how hard it is for information to change presuppositions.

There is one good reason why this myth survives and that is that up to now we have had no good means of measuring the interaction of stimulus and response in ordinary communications situations. We have had no way of measuring how much communication input produces what communication output. The

---

*Bernard Berelson, Paul Lazarsfeld, and William McPhee, Voting (Chicago: University of Chicago Press, 1954).

most elementary kind of effectiveness measurement has been unavailable. It is that problem which Dr. Kramer addresses in this monograph.

In this preface, let us look first at what the the special problems of measurement are that make it hard to estimate cause-effect relations in communication situations. Then we shall consider the resulting myth that exaggerates the power of psychological operations and the countermyth that denigrates them. Finally, let us note the possibilities offered for making better estimates of the quantitative relationship between stimulus and response by simulation of mass media processes.

Measuring Stimulus-Response Relations

In a psychological laboratory one can measure how far varying the amounts of what is communicated varies the response of the subject -- usually a college student hired for the job. To cite just one example among hundreds in the large literature of experimental social psychology, Whittaker has shown that the largest amount of attitude change occurred in his subjects when the propaganda they received was neither very similar to what they already believed, nor very different, but intermediate.* In the laboratory one can establish such quantitative relationships of cause and effect. One can measure the quantity and the

---

*James O. Whittaker, "Attitude Change and Communication-Attitude Discrepancy," Journal of Social Psychology, 65 (1965), pp. 141-147.

characteristics of the communication to which the subject is exposed and, by psychometric devices, the attitude or cognitions of the subject before and after. The finding of such experiments, as Hovland has noted,* is almost always that some measurable change occurs as a consequence of communication.

In the field, on the other hand, as Hovland pointed out, the usual finding is one of no attributable change following communication. Often from the before-measure to the after-measure there is no change at all. More often the world has changed between the two observations and so have some attitudes, but not in a way readily attributable to the stimulus communication. Too many intervening variables block change or make it random to permit any causal conclusion to be drawn. Among the intervening variables are the processes studied by Lazarsfeld and emphasized by Hovland in his classic essay comparing laboratory and field results, namely the freedom left to the audience members in the real world to select for attention only those communications that are compatible with their previous views and which will therefore not impel change in them. But the selective processes are not all encompassing. People do expose themselves to communications that produce change in them. Yet it has remained hard for the researcher to establish

---

*Carl I. Hovland, "Reconciling Conflicting Results Derived from Experimental and Survey Studies of Attitude Change," American Psychologist, 14 (1959), pp. 8-17.

conclusively the link between communications cause and attitudinal effect. Our techniques of observation have been inadequate to do that.

The social sciences have, however, developed excellent techniques for measuring attitudes and opinions of an audience and changes in them. The audience end of the communications process can be rigorously measured either in the laboratory or in the field. Public opinion polls, for example, spot changes of a few percentage points in what the public thinks. And because we have a good technique for documenting the response end of the continuum, communication research has tended to become audience research. We rarely study the full act of communication, which is a bilateral process between message and audience.

The original Cincinnati study which Dr. Kramer examines is an illustration of the point. The United Nations Association planned a major information campaign for Cincinnati. Wisely, they decided to evaluate the effort. They employed the National Opinion Research Center to do a survey of opinions and attitudes about the U.N. before and after the campaign. With many qualifications, the best one sentence summary of the results is that little planned change occurred. Attitudes and information about the U.N. were substantially similar after to what they had been before.

Note, however, the research design of the Cincinnati study. There were careful measurements of audience attitudes

before and after, but no measures of the totality of the stimuli that impinged upon the audience. The dimensions of what the campaign managers themselves put out were known, but not the much more massive product of the normal news media. It was assumed that the campaign was sufficiently massive that somehow it should have gotten through to everyone.

Doubt about the proposition that the entire public was effectively reached lay behind the Kramer study. We know that people select for attention a small percentage of the foreign affairs news that appears in their papers (13 to 22% is the estimate Kramer uses).* We know, furthermore, that no medium, much less message within it, gets to more than a fraction of the total audience. Knowledge of these facts stirred the suspicion that perhaps the reason that the observed audience changes during the Cincinnati U.N. campaign were so small was that the selective processes operating on the U.N. information output attenuated the frequency with which it actually reached Cincinnati citizens to a very low level. Perhaps it was a self-delusion of those involved in the vigorous activities of the U.N. information campaign to believe that it must be visible to everyone. Dr. Kramer's monograph tends to sustain the suspicion that frequency of exposure was not great. For example, comparing Kramer's content analysis figures with cumulative exposure figures that result from the simulation, we find

---

*See pp. 250-251, below.

that the average man saw or heard between 3 and 5 per cent of the messages put out and the average woman between 1 and 2-1/2 per cent. Thus even though there were 142 stories about the campaign sponsors, they reached the average man but seven times in six months and the average woman four times. Explanations of the U.N. appeared 43 times in the course of the campaign. The average man saw two of these, and the average woman one. There were 256 stories about the veto power, enough one might think to affect the attitude of any population, but the average man was reached by such stories only nine times in six months and the average woman six times. A news story or other tidbit on the veto once every three weeks or month is not going to change many attitudes in six months time if ever.

To arrive at estimates such as those we have just been quoting, one must first have means for measuring what stimuli actually reached the average audience member. We have for a long time had the capability of measuring propaganda output. Content analysis of media enables us to say how many column inches in newspapers, how many words in magazines, how many minutes on radio or TV is devoted to any given topic such as the U.N. Content analyses, however, have generally been divorced from audience studies. They have therefore had little to tell us about the cause-effect or communications impact process. They may have other values.

For example, a remarkable recent content analysis of news about the United Nations in press and broadcasts in 50 countries, directed by Dr. Alexander Szalai of UNITAR, revealed dramatically that the main U.N. coverage in 1968 (as Dr. Kramer found in Cincinnati in 1947) was of crises and international conflict, with but little attention to the developmental, welfare, and other day-to-day activities of the organization. However, such findings have in the past seldom been linked to measurements of the amount of attitude change resulting from the greater or lesser coverage received by different topics.

To make such a linkage of communication stimulus to audience response requires a more complex design than simply to correlate content analysis measures with attitude change measures. The reason for not being satisfied with such a design is that appearance of a column inch or minute of material in the media is, as already noted, not a good measure of opportunity for audience exposure. Appearance of a column inch at the top of page one of a newspaper with a circulation of one million is obviously not the same thing as a column inch buried inside a minor local paper. There are many factors besides mere content volume that must be taken into account if

one wishes to measure what stimulus material reached particular persons. For example, two inches or two items one after another in the same paper is not the same stimulus as one each in two different media. The same N-thousand people see both items in one case, more people see less material apiece in the other. Thus a simple correlation of content analysis results with measures of attitude change is most likely to produce but small explanations of variance, although with a strong suspicion that the weak relations are the result not of reality but of having only a poor measure of what actually came through to the audience.

Because content analysis does not measure what comes through to the audience, most researchers interested in measuring media impact have used interviews or similar audience measures not only for measuring attitude change but also for measuring exposure. Magazines and newspapers do not rely on their circulation figures alone to prove their reach, but commission audience surveys to measure readership among different kinds of persons and by kind of page. Radio and TV researchers use diaries, recording attachments, simultaneous telephone interviews, and recall to establish detailed ratings of shows at different times of day and by different kinds of people.

For at least three reasons, however, even this data, excellent as much of it is, does not meet our need for an independent measure of exposure to relate to observed responses.

First, and probably least important, is the fact that measuring both exposure and response by interview answers may result in confounding of that which we wish to separate. A person's statement about what arguments or messages he read or heard is in part a statement about which arguments or messages persuaded or impressed him, rather than being a pure measure of exposure. That problem is rather acute in studies (such as the Cincinnati one) which rely for all their data on social attitude surveys or public opinion polls that also are used to measure respondent attitude. It is less acute for the rating services and similar media studies which use somewhat carefully controlled measures of exposures such as a dummy magazine for aided recall or a diary.

A second, more serious, problem is that carefully collected rating data is often unavailable. It is usually unavailable to the social science researcher who cannot afford a careful field study of exposure to the particular output (such as the U.N. campaign) in which he is interested. He must usually be satisfied with standard statistics of circulation and audience not pinpointed to his particular stimulus material. From those general media statistics he must calculate estimates of the audience for the material of particular interest to him. That is one of the things our mass media simulation facilitates and which Dr. Kramer did in the present study.

Furthermore, for many purposes, such as studies of the reach of communications in closed societies, the researcher is not able to collect data in the field. Dr. Kramer's study, done in the United States where abundant audience data was available, served as a methodological test for a simulation that in other studies is being applied to measuring communication in the Soviet Union.

A third problem is that exposure data collected by interview or other audience interrogation devices is usually collected one medium at a time. Nielsen ratings exist for TV shows. Each major magazine or newspaper has its own audience study. It is more than a trivial matter to put these together. Suppose we know that 20 million persons saw a particular TV show with a message of interest and that 20 million saw a magazine that contained the same message. One cannot conclude that 40 million were reached, since some persons saw both the TV show and the magazine. What can we conclude about how many persons are exposed once, twice, or zero times? As Dr. Kramer explains in this monograph, our simulation model is designed in part to produce estimates of that kind.

All the considerations just outlined as to why neither content analysis, nor audience surveys, nor ratings alone give us an adequate measure of exposure led us to design the mass media message diffusion simulation that is described and tested in this volume. The purpose of the simulation program is to provide estimates of who was reached how often by

what messages in a national or other population subjected to some information flow. The inputs are data about factors which must be taken into account, including the volume, frequency, formats, and location of messages, and the size and characteristics of the audience population, their media habits, and special events (such as crises) in the real world which may change their media habits. The simulation accepts whatever information there is about all these variables and then puts out estimates of what kinds of persons have been reached, how often, by each theme in the contents. That is a stimulus measure that is, hopefully, more useful than the crude content analysis with which we start, and therefore one that is more appropriate to correlate with attitudinal response data.

That is what Dr. Kramer has done for the Cincinnati U.N. campaign. Why that campaign? That is a rare case where there exists well collected before and after attitude measurements, and also where there was a well defined propaganda campaign that could be documented by a content analysis, and finally where all the necessary population and media data exist so the content analysis results and that data could be fed into the simulation model. The Cincinnati U.N. campaign was therefore an ideal instance on which to make a first trial of the mass media message diffusion simulation, to see how it worked and whether it produces any insight as to the impact of the flow of messages on attitude change.

The goal toward which we are working, one toward

which some steps have been taken in this first test, is to create a device that will permit us to do studies in nature such as those we now do only in the laboratory, that measure how much impact propaganda or information have on the people who receive it.

How Much Impact Does Communication Have?

We alluded above to the myth of all-powerful propaganda. It stems in part from the claims of advertising men and public relations men themselves, for they need to emphasize the value of their services. It also stems in part from World War I and the argument of German nationalists that they were not defeated on the field of battle, but were taken in by Allied propaganda. More careful studies of psychological warfare experience give little support to the notion that the airdropped leaflets had any major impact in changing German minds. What every careful study has shown is that well timed, well planned psychological warfare can be very effective indeed, if -- but only if -- it provides information to help the audience do what it is already predisposed to do.* Janowitz and Shils in a classic study found in World War II that Wehrmacht members were likely to respond to surrender appeals only to the extent that cohesion had already broken up in their

---

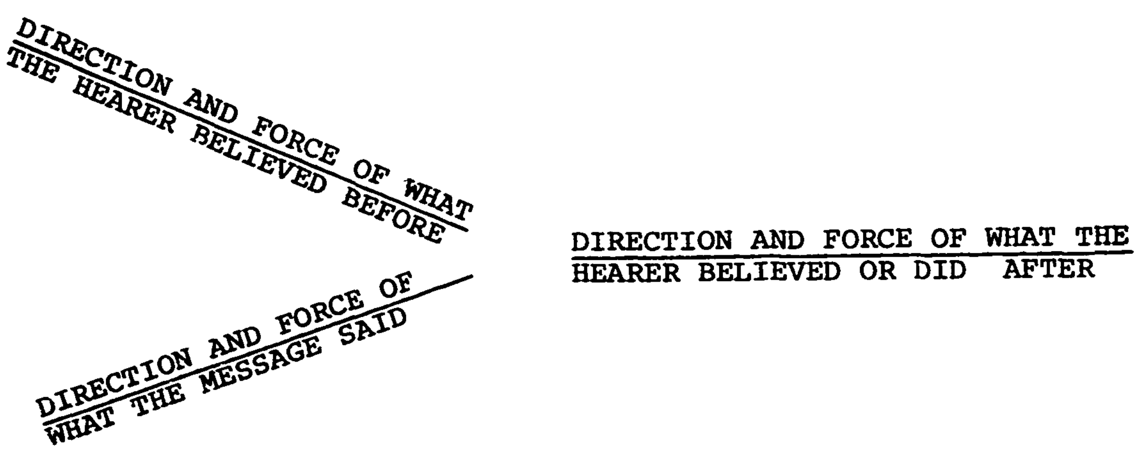
*Daniel Lerner, Sykewar (New York: G.W. Stewart, 1949).

small units by the failure of leadership by the non-coms.*
In Viet Nam studies of Viet Cong returnees to the government side showed that by 1966-67 virtually all Viet Cong members knew about the Open Arms program through airdrop leaflets and radio broadcasts and knew from these how to go about surrendering, but that actual defection occurred only when family pressures, or appeals by former buddies, or mistreatment in their Viet Cong unit created a first-hand personal pressure.

Analogous points could be made from the literature on advertising, or the literature on election campaigns, or the literature on adoption of innovations. All of them show no simple direct relation between the volume of mass communication and the audience response.

There are two functions which mass communications do serve in the complex process of producing an attitude change or action. Mass communications lay a foundation of basic information, creating in that way the environment in which any other influences may act, as, for example, enabling the Viet Cong to know how to surrender if they wish. Secondly, mass communications or any other kind of communications may be the trigger which sets off a reaction the character of which is determined far more by the system that is triggered than by the trigger itself.

In the literature on communications effects two main models may be found, the so-called vector model and the so-called trigger model.* The vector model sees the audience's predisposition as a force going in one direction and the communications input as another force going in another direction, and the net effect in behavior as a resultant of these two forces.



That vector model is not necessarily naive. It does not treat propaganda as all-powerful, as do some popular writers on persuasion. The model allows appropriate weight to the predispositions of the audience. It takes account of the inertia against change rooted in the hearer's prejudice and made effective by mechanisms of psychological defense. The vector model also encompasses findings as subtle as the "sleeper effect," in which the fact that messages come from disapproved sources is gradually forgotten, while their contents come to be accepted. The vector

---

*Raymond A. Bauer, Ithiel de Sola Pool, Lewis Dexter, American Business and Public Policy, New York: Atherton Press, 1963, pp. 466-472.

pattern which expressed that finding shifts through time, so that vectors which were earlier opposed to each other come to be seen as having the same direction.

But many phenomena that do not fit the vector model do fit a trigger model. A trigger is something that starts a process, but has little to do with how the process works. To analyze the trajectory of a bullet one needs to know the charge behind it, its shape, the bore of the gun, the direction of aim, but not anything about the trigger. The trigger only sets off a process whose determinants were in the things it touched. Communications can work that way.

> First, there lies latent in individuals a great collection of traces of previous communications. Any new communication may serve to change this massive structure only imperceptibly, but it may at the same time set it into action in directions determined by the structure itself more than by the trigger stimulus.
>
> Second, the event triggered within the system may itself have more effect on the system than does the original stimulus. Arguing for one's views in reply to a challenge may have more effect on one than does the challenge.
>
> Third, within an individual many latent attitudes may be simultaneously present. The structure of social controls and social relations may make some of them easier to express than others. Thus, even stimuli which have a persuasive effect on a man's thought may trigger quite opposite expressions.
>
> Fourth, where a stimulus is addressed to a population of individuals, structural determinants may result in its mobilizing different proportions of those who agree and those who disagree with it. If it mobilizes more of those who disagree, the stimulus may boomerang.*

---

*Ibid., pp. 469-470.

Examples of such trigger effects were found in the foreign trade debate of 1960-61 where pro-free trade propaganda stirred up protectionists to action more than it did those who sympathized with the propaganda.

If all the points just made about the complexity of the cause-effect relation between communication and response are correct, then why should we be trying to create a device to permit estimates of correlation between quantity of exposure and quantity of attitude change? There is every reason to expect the correlations to be small. Communications will only work on those groups and in those situations where predispositions are favorable. For long periods massive information may only be setting a latent environment of knowledge and images. In other periods a small stimulus may trigger off that mass of latent predispositions.

Yet even though our expectation is for weak relationships, it is of value to create a tool that enables us to measure more precisely than we now can what the magnitude of these relationships may be. In the first place, one cannot dismiss the vector model. There is substantial evidence from laboratory experiments and from cost-effectiveness studies of advertising that repetition has its effect. At least under some circumstances, the amount of persuasion is a function of the amount of communication. It would be useful to know how often, how much, and under what circumstances that is the case.

Secondly, even in trigger situations there are

quantitative parameters. How many people get exposed enough for the trigger to go off? What kinds of people are these? How much latent information had to be there first, and when and to whom did it come?

Thus Dr. Kramer approaches the simulation model neither with the naive expectation that there should be a powerful relation between the amount of exposure to U.N. information and the amount of attitude change, nor with the superficial notion that nothing can be learned by examining the quantitative relations between them. The mass media simulation permits us to learn more about what the dynamics of the relationship are.

The Simulation

We have already said enough about the mass media simulation that is exercised in this volume so that the reader knows that its purpose is to provide an estimate of the frequency of exposure of different types of audience members to different themes in the media content. At first glance it is an extremely naive simulation. It represents in the computer a sample of the population each member of which has certain demographic characteristics and each member of which also has certain media habits, such as subscriptions or probabilities of turning on certain broadcasts. Then a scenario is written which locates each occurrence of each theme in each issue of each medium. Then the computer passes the scenario over each individual in

the imaginary sample and keeps a count of each probable exposure.

If there were no other problems to it, that would be a very pedestrian, inefficient number grinder. These operations could be done better by solution of probability equations. There are, however, four subtle problems that enter the simulation, that give it its intellectual interest, and that make mandatory a simulation approach rather than an analytic solution. These are all discussed in Dr. Kramer's volume. They are designated by the labels cumulation, duplication, triggering, and estimation of full arrays from incomplete information.

Cumulation: Messages in the mass media are events in vehicles that appear in regular patterns over time. Broadcasts are patterned into daily or weekly programs on discrete channels. Print media appear as serial issues of fixed publications. The members of the audience have regular habits regarding these broadcasts or publications. The significance of that fact is that message exposures are not statistically independent events. If John Doe subscribes to a pro-U.N. paper then the chance of his reading repeated stories about the U.N. is high. It will not do to randomly assign exposures to stories about the U.N. to all members of the hypothetical population. Some of them have media habits that will expose them often, and some have habits that will expose them seldom. Consider further the difference between a scientific journal and an airdrop

leaflet. Virtually all issues of a scientific journal go to subscribers so that the probability of a person who sees one issue also seeing the next is perhaps 60% or 70% or 80%. An airdrop leaflet, on the other hand, is received pretty randomly. The chances of seeing two in a row may be small. Consider the implications of this. Consider two media both of which reach .5 of the population. In two issues, a medium which always goes to exactly the same persons would reach .5 of the population, each exactly twice. Another whose distribution was completely random would have reached .75 of the population; .5 of the population once, .25 twice. Dr. Kramer developed the problem of cumulation and a model for dealing with it in this volume. It is a problem which has been much discussed in the research literature on advertising. Dr. Kramer's treatment raises the discussion to a new level of sophistication.

Duplication: Cumulation is a problem because exposure to messages are not independent events but are autocorrelated over time. Duplication is a different but analogous problem. Exposure to different message vehicles is not independent. They are correlated with each other. For example, a person who reads the Atlantic is more likely than the average person to read the New Yorker, and less likely to watch Gunsmoke. But is the Atlantic reader more or less likely to read Harper's? The answer is not obvious. Perhaps the Atlantic saturates most of its readers' desire for elevated current commentary or perhaps

that factor is overridden by the similar taste to which the two magazines appeal. Whether it turns out in fact that these magazines are primarily surrogates for each other or share a joint demand, it is nonetheless clear that they are not statistically independent. Somehow account must be taken of the fact that use of one medium has implications for use of the other.

Triggering: The scenarios that we write in the simulation serve to map out the occurrence of various messages. These messages reach people as a function of their basic media habits. But the messages may in turn modify these habits. If a person is exposed to the message "Soviet troops invade Czechoslovakia," he is likely to rush to his radio set and keep it on, hearing much more news than he otherwise would have. The simulation allows for such triggering by which normal media habits get changed. Note that in the simulation we use the word "triggering" in a sense that is a special case of what we above called the trigger model of communication. We use the term triggering to describe one particular response to a communication trigger, namely a change in media habits.

Estimation of Full Arrays from Incomplete Information: By now it will have occurred to the reader that the simulation Dr. Kramer is using calls for a vast amount of information. We need information about the demographic characteristics of the population, their habits with regard to every medium, the details of occurrence of every message, cumulation data, duplication data, and a few other things. Clearly all that data never exists.

One of the main design objectives for this simulation was that it should accept whatever data does exist and estimate from that whatever data is lacking. For that operation it was particularly important that the data input phase of the simulation be interactive between researcher and computer, for during data input and estimation there may be many decisions to be made.

Chapter II of Dr. Kramer's monograph, especially, describes some of the main routines for estimation of missing data. We rely heavily on a routine which we have named "Mostellerization" after the scholar whose idea we use, and which is designed to provide estimates of cell values in a matrix where we know the marginals and also have some basis for specifying interactions.

Results

While the main value of Dr. Kramer's work is clearly methodological, it does tell us some interesting things about the much-studied Cincinnati U.N. campaign. It confirms the suspicion that except for crisis news, despite the massive efforts of the United Nations' supporters in Cincinnati, many citizens heard very little about the U.N. during the campaign.

The study also shows a rather clear correlation between the themes most covered in the news and the themes on which most change occurred over the period studied. The period was one of much international crisis activity and during it fear of war increased markedly. On the other hand, routine U.N. information got through to the public very little and attitudes

and information about normal U.N. roles changed but little. Thus, across themes we find a substantial correlation between the messages carried and the audience response.

The evidence that public opinion changed most in regard to those topics on which it was most exposed to mass media messages is presented in the following table, which matches up as well as we can topics covered in the mass media with questions asked on the NORC survey. Clearly exposure to information was a prerequisite for change in attitude or knowledge.

On the other hand, Dr. Kramer's results show only limited correlations across audience types between amount of exposure and amount of attitude and information change. It is only partially true that the groups most exposed changed most.

How can we reconcile these superficially contradictory results? To some degree it must be recognized that some flaws in the research design may be involved. Data on radio broadcast content from 1947 no longer existed in 1968. Dr. Kramer therefore postulated that radio news flashes covered the same topics as did the newspaper front pages. Thus even if there were significant differences between the radio and newspaper audiences, differences in what they received in the simulation would be attenuated by the similarity of the scenario offered them. What could appear in the simulation results would only be differences in what reached the audiences of the different

Themes in order of Frequency	Number of times counted in content analysis	Per Cent Changing Attitude or Knowledge in NORC Panel*	
3. Violence, threats to peace, war	2202	Expect war within 10 years	.228
		War main problem facing U.S.	.213
2. Russian-American relations	752	USSR main problem facing U.S.	.191
1. U.N. peacekeeping	595	More interested in U.N. because situation more critical	.209
4. Great Power discussion in U.N.	357	Pessimistic that U.N. will succeed despite disagreements	.177
5. 6. 7. Veto	234	Knowledge of veto	.090
9. Campaign sponsors	142	Know names of campaign sponsors	.059
11. 12. Satisfaction, dissatisfaction with U.N.	138	Dissatisfied with U.N.	.027
10. Explanations of U.N.	43	Knowledge of various aspects of U.N. (average over 6 items)	.050
8. U.N. and human rights	32	U.N. job to guarantee equal rights	.038

*Calculated from pp. 410 ff., below.

newspapers and between the reach of the kind of major story carried on the front page and the kind of deeper background material carried inside. Thus variations in messages received that might correlate with attitude change across groups are weaker than we would like them to be in an ideal test and weaker than they have been in other cases to which we have applied the simulation.

Even if that had not been true, however, there would still be reasons why across-groups correlations would come out weak, reasons that may well have reduced whatever residual correlations the data structures would have allowed. Let us ask ourselves what it is that gives impact to a theme. Is it its absolute frequency or is it its frequency as a percent of messages received? Consider the comparison of an educated middle class person who reads and listens to much news with an uneducated lower class person who has little interest in public affairs. The former is apt to have rather well formed political opinions; the latter rather amorphous ones. The former reads and hears much more news of all sorts. In that total there will be more items about the U.N. But these presumably have no more impact on the well formed views of this well read man than the few items about the U.N. that reach the more malleable cognitive structure of the less informed man. Indeed, if we apply the findings of such studies as Voting,*

*Op. cit.

we might expect even more changeability among the less informed than among the well informed. That, indeed, is what we do find in this study. Dr. Kramer correlated citizen exposure to each theme via the mass media with growth of information about foreign affairs between the two waves of the NORC panel survey. (The correlation was an ecological correlation across respondent types.) He used two measures of increased information: large increase of information, i.e. 3 to 5 more correct answers given to the five questions asked at the end of the six months than at the beginning, and moderate increases of information, i.e. 1 or 2 more correct answers given at the end than at the beginning. Clearly to have recorded a large increase in knowledge a respondent must have started out quite ignorant, at very best getting two out of five questions right. It follows that the audience types among whom large increments of knowledge occurred were audience types that were not much involved in foreign affairs. They consisted of people who knew little because they read little. On the other hand, moderate increases of information on the five question test could be the result of high exposure to media messages.

That explains what at first glance may be a puzzling finding in Dr. Kramer's results. There is a <u>negative</u> correlation of .7 across audience types between exposure to the media and <u>large</u> increases of information and a <u>positive</u> correlation of .7 between exposure and <u>moderate</u> increases of information. That is as it should be. Furthermore, by a small amount, the

latter correlation was highest for those themes most fully covered in the media: war and peace, and US-USSR relations. That is also the direction it should be.

Those results provide support for the realism of the simulation. So do a few other results of Dr. Kramer's validation effort. Those persons who habitually use radio for news are by all evidence less informed and less interested in world affairs than those who use newspapers and meetings. Exposure to various themes in the simulation correlated negatively with interview mention of radio as a news source and positively with mention of newspapers and meetings.

We note also that the six month period of the interview study was one of growing dissatisfaction with the U.N. due to international conflict and growing fear of war. The extent of dissatisfaction was, however, negatively correlated with exposure to the news. Better informed people were less volatile in discouragement. In this instance differential exposure to different themes makes a clear difference. There is only a -.3 correlation between amount of exposure to the general news coverage of war, peace, and Soviet-American conflict and dissatisfaction about the process of the U.N., but there is a -.6 correlation between such dissatisfaction and exposure to such matters as explicit explanation of the U.N.

Finally, Dr. Kramer finds a +.5 to +.6 correlation between exposure to the various media messages and increased knowledge about the U.N. veto, but substantially no correlation

between media exposure and knowledge about such other U.N. matters as human rights, health, or international trade activities. It will be recalled that the simulation provided estimates that the average male received about nine news messages on the veto in the six months. The corresponding figure for the U.N. and human rights is one message in six months and for the U.N. and health and trade, to the nearest approximation, zero. Thus the simulation results are once more supported as reasonable by data from the real world. Attitude changes appeared in the NORC interviews where the estimated amount of exposure to a theme reached a critical, even if quite modest, level. Where the exposures were minimal, the attitude changes were also minimal or random.

Thus even those negative results which Dr. Kramer found in his validation effort are on second thought plausible. It is plausible that themes which get so little attention that they disappear in the noise of a mass media system have no measurable impact on attitudes. It is also plausible that a mass media system which delivers substantially the same vehicles of information to the whole public, will produce very similar correlations across groups between attitude changes and the frequency of exposure to any particular theme. To a substantial extent exposure to any one theme was an index to exposure to the highly uniform mass media system as a whole, with all the themes it contained. Furthermore, personal influence and word of mouth communication, which are not represented in our

model, keep the society bound together in relatively high cohesion. If real events cause changes in the expectation of war among those who follow the news closely, these new expectations will be carried to the less attentive citizens from the more attentive ones by word of mouth.

We are not surprised, therefore, that in a relatively homogeneous society like that of the United States we find that the news focus does determine the issues about the U.N. on which change takes place and those on which it does not, but that changes when they do take place at all, take place fairly uniformally across broadly defined demographic groups.* Dr. Kramer's findings in this respect are highly plausible.

The validation of a complex simulation model is not easy and cannot be done by any one set of observations. The first test of the validity of a complex model is the face plausibility of its output. More advanced validity tests are prediction or postdiction in situations where the models' parameters can be matched to the real world; does the simulation output follow a distribution that is compatible with our knowledge of the real world? But a complex simulation must

---

*This finding matches those in a study of changing attitudes towards violence done by this author for the National Commission on the Causes and Prevention of Violence, "Trends in Public Opinion About Violence: 1937-1968," November 1968, and one on attitudes toward the Viet Nam war done for the United Nations Association, "Trends in American Public Opinion on Viet Nam," March 1969.

match the world on many observations before we can give it much credibility. The more variables there are in the model, the more test observations are required. And there is no fixed point at which the model is proven. Validation is a matter of degree. Dr. Kramer ran a simulation of the Cincinnati U.N. campaign because it is a simulation of a situation on which the real world data was relatively rich. The results of the simulation of that situation carry us the first steps in validation of the simulation model. The results obtained so far are all at peace with the data that exist from 1947. Certainly we cannot argue that validation has been carried very far by this one set of tests. It has, however, been begun. Dr. Kramer has shown extraordinary care and methodological dexterity, both in his contributions to the development of the model, most of which was specified and programmed by him, and in this initial testing of it.

Ithiel de Sola Pool
Center for International Studies
Massachusetts Institute of Technology
July 1969

ABSTRACT

A Computer Simulation of Audience Exposure In A Mass Media System: The United Nations Information Campaign in Cincinnati, 1947-1948, by John Francis Kramer

During six months of 1947-48, several organizations in Cincinnati conducted a public information campaign to educate the population of Cincinnati about the United Nations. An NORC panel survey which was conducted to study the effects of the campaign found very slight changes in information and opinions about the U.N., but large changes in concern about war and relations with the U.S.S.R. To better understand these effects, we have used a simulation of the Cincinnati mass media system to predict the frequency and reach of the flow of messages in the system from known facts about the population, its media habits, and the messages in the mass media.

Data about the population and its media habits were taken from census statistics and newspaper and radio audience studies. A content analysis of the press provided a set of messages relevant to the attitudes and opinions measured by the NORC survey; radio messages were postulated from the newspaper messages. The themes of messages thus identified by the

content analysis served as input to the simulation.

Before running the actual scenario, we ran three trial scenarios to test the plausibility of the resulting exposures and the consistency with which the model synthesized the input data for the media system. We found that the data were consistently synthesized and that the exposure output was plausible.

For the twelve real themes we described by sex, education, and socio-economic status, the growth and distribution of exposure, the most and least exposed population subgroups, the duplication of exposure across themes, and the most important media types and vehicles in producing exposure to these themes. We found that those themes which seemed most closely related to the large changes in the NORC panel were also those themes for which the simulation predicted the highest average frequency of exposure for the population. Continuing the analysis, we proposed two basic models relating information or attitude change to frequency of exposure and then attempted to correlate, across sixteen population subgroups, the predicted exposure to the themes with changes in attitudes and information and recall of exposure in the NORC panel. We found that the distribution of exposure across the population subgroups was relatively constant from theme to theme and that the resulting correlations were small and not easily interpreted. We concluded

that any effects of the mass media themes in the panel at the subgroup level were probably confounded by variations in prior exposure and exposure via informal communications channels. Finally, we offered several specific suggestions for increasing the usefulness of the model by improving the data base and modifying the output statistics and organization.

ACKNOWLEDGEMENTS

The computer model here described is the product of several minds; however, the programming of the model fell entirely upon a friend and fellow graduate student, Herbert J. Selesnick, and myself. Throughout countless hours, we programmed the model together and enjoyed each other's company. I am indebted to several of the faculty of the M.I.T. Political Science Department and especially to Professor Ithiel de Sola Pool. He has not only guided and advised the preparation of this work, but has also taken a continuing interest in my work as a graduate student at M.I.T. as a friend as well as a teacher. Without his guidance and support, this monograph would not have been possible.

Neither H.D.W. nor I may ever know how important was her help and consultation during difficult periods while working on this piece; I offer her my deepest appreciation. And to J.B.S. who preserved my well-being with her love and enthusiasm, my continuing love and respect.

The research for this monograph was supported in part by a grant under contract number 920F-9717 of the Advanced Research Projects Agency, Department of Defense,

administered through the Center for International Studies at M.I.T., and in part by Project MAC, an M.I.T. research program also sponsored by the Advanced Research Project Agency under Office of Naval Research contract number Nonr-4102(01). In addition, the Political Science Department and the Computer Institute for Social Science Research (CISSR) of Michigan State University generously supported the final computing runs and drafting of my thesis.

TABLE OF CONTENTS

Preceding page blank

Chapter | Page

LIST OF TABLES

Table		Page

Table		Page

LIST OF ILLUSTRATIONS

Figure		Page

Figure | Page

Figure		Page

Figure		Page

INTRODUCTION

THE GENERAL MASS MEDIA SYSTEM

A mass media system in a typical city might be composed of several newspapers, several television stations, five to ten radio stations, magazines, transit advertising, billboards, etc. Surely the number of exposures of individuals to messages carried in the mass media is many times as large as the population. Often it is conceivable that most of the media are carrying a story, a message, or a common theme simultaneously. How is it, then, that great proportions of a population hardly ever become acquainted with some of the themes carried by the mass media? With all the different audiences characteristic of the various vehicles in the mass media, how can one understand and predict which segments of the population will be exposed to given themes carried in certain parts of the mass media? These are the questions which we have posed in this thesis and which we attempt to answer by using a computer model of the mass media system. The simulation can bring order out of the chaos of, at a minimum, tens of different audiences to the various newspapers, radio broadcasts, television broadcasts, magazines, etc. We will attempt to show that

several assumptions plus the integration of several kinds of data can go a long way toward explaining why certain members and groups of the population are repeatedly exposed to themes in the mass media while other groups are rarely, if ever, exposed.

The Possibility of a Mass Media Simulation

Students of public opinion from political scientists to Madison Avenue advertising executives have long wished to explain and understand the flow of messages and information in a given public or audience. Many studies have been performed since the beginning of World War II on elements of this system of information and message flow. Some of these studies have focused on the channels of communication, contrasting formal and informal channels, the mass media versus word-of-mouth communication. Other studies have emphasized the number and content of the messages in these systems and the factors which control them. Still other studies have attempted to describe the audiences of certain channels and for given messages and the effects which the messages have on these audiences. Occasionally someone would study a combination of these factors and the interactions between them. These studies have provided empirically based insights into the workings of various elements of the total communications systems for a public, but until very recently we have not had the means to integrate all the

information produced into a description of the total working of a communication system. However with the advent of high speed, large capacity computers it has become feasible to attempt an integration of many of the findings of previous studies into a simulation of a rather large part of the communication system of a given population.

In 1961 and 1962, the Simulmatics Corporation developed an early model of a mass media system which processed data about the average audiences of various communications vehicles and the (real or hypothetical) messages programmed over these vehicles, and reported exposures in various classes of the population and their frequency and cost.[1] This early model was developed under the direction of Professor Ithiel de Sola Pool of MIT. In April of 1963 the Advanced Research Projects Agency (ARPA) of the Defense Department agreed to fund at the Center for International Studies at MIT a program of research on communication in several European and Asian communist countries under the direction of Professor Pool. Included in this project was an attempt to build a computer simulation of a media system which might integrate many of the various kinds of aggregate data gathered in the research project and attempt to predict from this aggregate data on audience

---

[1]The model is reported in a pamphlet by the Simulmatics Corporation entitled Simulmatics Media Mix: Technical Description (New York: The Simulmatics Corporation, 1962).

behavior and from the location, content, and format of messages in a media system such variables as the types of people exposed with what frequency, the reach or coverage of a given message or theme of messages, the time required for the spread of the message or the rate of penetration of a theme or message in the population, the duplication between the various kinds of messages or themes, etc. From that time to the present many social scientists and graduate students have worked in some degree at implementing the simulation part of this communications project. Two of the algorithms used in the present simulation model bear the names of Professor Frederick Mosteller of Harvard University and Professor Robert Abelson of Yale University who suggested solutions to rather knotty computational problems which arose in developing the model. The responsibility for the final programming and implementation of the model and the integration of many of the tentative solutions proposed by the various people working on the model over the years, plus solutions to new problems which had not been foreseen and a complete restructuring of the order of the simulation and the method of calculation of the outputs of the simulation (which was necessary to make it indeed feasible to run such a simulation) devolved upon two graduate students at MIT during 1966 and 1967 under the direction of Professor Pool, John F. Kramer and Herbert J. Selesnick. This thesis is an attempt at a partial validation of the resulting model.

There are several different measures of the validity of a model (e.g. its "face" plausibility, the amount of its unaccounted for variability, etc.), but one of the most stringent and therefore convincing measures (when successful) is the comparison of the predictions of the model with empirical data. In order to so test the simulation, we needed data which would enable us to describe a mass media system in terms of the parameters of the model, to locate and analyze a set of messages in that system over a specified time interval, and to measure the exposure of the population to these messages during that interval. If this last criterion was impossible to meet, we could measure the population changes in relevant attitudes and information during the interval and attempt to correlate these with exposures predicted by the simulation.

A search for a situation in which all three kinds of data were present lead us to the well-known 1947-48 U.N. information campaign in Cincinnati. Concurren' with th. campaign, the National Opinion Research Center (NORC) conducted in Cincinnati several surveys of attitudes and information relating to the United Nations and other international news. With these data in hand we looked for the other required data and for the most part have been able to piece together enough data from many sources to describe the mass media system and the flow of relevant messages in it during the six months of the Cincinnati campaign.

Examining closely the changes reported in the panel by the NORC data, we found that the greatest changes occurred not in opinions and information relating to the U.N., but in the areas of expectations of war, relations with Russia, and control of the atomic bomb. For this reason, a substantial portion of the research was in these areas, with the goal of correlating the amount of change measured in various subgroups of the Cincinnati population with exposure of the subgroups to relevant messages as predicted by the mass media simulation.

Chapter I presents a brief summary of the Cincinnati campaign and a description of the simulation model. In Chapter II, we look in some detail at one of the most important routines in the simulation, the program for estimating several model parameters from incomplete data. The next five chapters outline the many levels of analyzing and organizing the input data for the simulation. Chapters III-VI describe the construction of the mass media system, i.e. the specification of the model to the Cincinnati population and mass media vehicles as they were structured in 1947-48. The content analysis and description for input to the simulation of the actual messages during the campaign are presented in Chapter VII. Finally in Chapter VIII we discuss the simulation output from six trial and twelve real themes of messages. We find that the simulation does consistently synthesize and reproduce the input data and

that it produces plausible exposure values for the real themes of messages. We conclude with an evaluation of the simulation attempt and several suggestions concerning the strategy of simulation.

The simulation model was originally programmed entirely for the Project MAC time-sharing system on the IBM 7094 computer. Interaction between the computer and the researcher is greatly facilitated by this system; often the researcher can type in his data, immediately view the consequences, and accordingly modify the data if desired The current simulation model allows up to 64 media vehicles, with a maximum simulated population of 5000 persons, distributed across a maximum of 400 population subgroups. The last link of the simulation, which performs the actual processing of the messages and recording of exposures, is presently being reprogrammed for the batch-processing IBM 360-65. The larger capacity of this machine and increased processing speed will allow more detail and easier handling of the output statistics.

CHAPTER I

SIMULATING EXPOSURE TO INTERNATIONAL AFFAIRS IN CINCINNATI

The Cincinnati U.N. Campaign

During the six-month period from September 1947 to March 1948, the American Association for the United Nations and the United Nations Association of Cincinnati conducted one of the best documented and most conspicuous failures of a mass educational campaign to be found in the literature of public opinion and propaganda.[1] "In an effort to stimulate interest and convey information about the United Nations [the] two organizations literally bombarded the city of Cincinnati"[2] with information in an intense campaign whose "objective was to reach in one way or another every adult among 1,155,703 residents in Cincinnati's retail trading zone."[3] The newspapers played up United

---

[1]Reported in Shirley A. Star and Helen MacGill Hughes, "Report on an Educational Campaign: The Cincinnati Plan for the United Nations," American Journal of Sociology, January 1950, pp. 389-400.

[2]Raymond A. Bauer and Alice H. Bauer, "America, Mass Society and Mass Media," The Journal of Social Issues, Vol. XVI, No. 3, 1960, p. 12.

[3]Star, op. cit., p. 39).

Nations news and information; for one week one hundred and fifty spot announcements were broadcast over one of the radio stations. "In all, 59,588 pieces of literature were distributed and 2,800 clubs were reached by speakers supplied by a speakers' bureau and by circular, hundreds of documentary films were shown . . ."[4]

The magnitude of the failure of the campaign to stimulate interest in the U.N. is well documented by two NORC surveys conducted at the beginning and at the end of six-months period.[5] With respect to the level of information "the before and after scores remained remarkably constant; for example, in September, 34 percent said they had heard of the United Nations' veto power and 7 percent could explain how it worked; in March these figures were almost unchanged--37 percent and 7 percent."[6]

During the six months, however, there was a violent conflict raging in Palestine between Arabs and Jews, and in

---

[4] Ibid., p. 392.

[5] These surveys include a panel of 745 persons (of which 20% were lost or refused to be reinterviewed) plus a second sample of 758 adults interviewed in March. The questions included measures of information, interest, attitude toward the United Nations and general international affairs news and for the March interviews measures of exposure to the information and the channels by which the respondents were exposed. Full details of the surveys appear in Cincinnati Looks at the United Nations and Cincinnati Looks Again, being Reports Nos. 37 and 37A of the National Opinion Research Center, University of Chicago, Chicago, 1948.

[6] Star, op. cit., p. 392.

the U.N. somewhat less violent conflict over the partition of Palestine. East and West were regularly denouncing each other, not only over the issue of Palestine, but also over the use or misuse of the United Nations in relation to the issue. We shall see in Chapter VII that these were the themes predominant in the mass media and it is therefore not surprising that the large changes in public opinion were in this area. For example, in September, 1947, 24 percent of the sample named war or peacekeeping as the most important problem facing the country; by March, 1948 this figure had risen to 45 percent. The percent in September expecting war in the next ten years was 49 percent; in March it was 72 percent.

Since our model is a model of the flow of messages in the mass media, and much of the effort of the information campaign was directed into less formal channels of communication which we did not model, we should not consider this simulation a replication of the total information campaign, but only of the relevant messages in the mass media, plus other international news messages which relate to the larger opinion changes observed in the Cincinnati panel. Before we look further into the data relevant to these changes, we will present the model of a mass media system used in this simulation.

<u>Strategies for Simulating a Mass Media System</u>

How shall we model a flow of messages in a mass media system? We might choose one of two strategies for building a system to describe the flow of messages. The first strategy would make relatively few strong assumptions based on research studies and intuition about the message flow to develop a general model, not completely specific to any given situation and not strongly tied to data from any particular mass media system. Such a model would be at a high level of abstraction and because of its generality it would probably not be a very accurate predictor of the message flow in any particular mass media system, but rather a heuristic device which would attempt to capture some essential part of the flow of messages across many media systems. A second way to approach the simulation of a mass media system is to make a model which incorporates all kinds of specific data to model a particular mass media system. In this kind of model the level of abstraction would be significantly lower and the heuristic value, or insights into the functioning of the system, would probably be less. but the ability to make predictions about a specific system would quite likely be enhanced because of the use of large masses of data from the particular system. In general in social science the following rule seems to be true: the higher the level of generality of a model, the less useful it is as a predictor in a particular situation.

This latter approach is the approach we have taken in constructing the present mass media simulation. We have attempted to design a model in a set of computer programs that will allow a researcher to use as much data of various kinds as he can muster pertaining to a particular mass media system, incorporating this data into the model to produce flows of message exposures specified throughout the population for the particular mass media system over time.

What kind of data might one have for a particular mass media system? Typically we find data about the audiences of the vehicles of the mass media system and how they distribute themselves throughout the population. By this we mean that the audiences are described in terms of their composition by sex, age, education, level of economic status, race, and perhaps by other demographic, social, political, or public opinion dimensions. These data come from surveys and studies often commissioned by the management of the media vehicles in a particular area. Usually, at least one of the newspapers in a metropolitan area will conduct a study to see what kinds of people are reached by the newspaper versus the kinds of people the competitors are reaching in order to better make its case for advertisers. Of course, the ratings for radio and television and the description of the audiences of these vehicles by demographic and social types is well known. Sometimes also, but very rarely for daily media vehicles, there are data

which describe the growth of the audience exposed at least once to the communication vehicle after several issues of the vehicle have appeared. This net coverage of the vehicle after a given number of issues we call the cumulation or cumulative audience of the vehicle after several issues. These data are particularly important because they give an idea of a differentiation within a population between people who are almost constantly exposed to the vehicle and people who are much less frequently exposed to the vehicle. In addition, we occasionally will know the average duplicated audience between some pairs of vehicles.

The data which we have described above represents the best possible situation which one actually encounters in attempting to describe empirically the mass media system. We note that these data are aggregated or macro-data of average statistics about groups in the population. Usually the data available for a system is much sparser than the data described above. For some or many of the vehicles, often not even the average audiences are known; these must be estimated from some other kind of information, perhaps by drawing analogies to other situations where such data are known. In constructing the simulation we have made the assumption that some macro-data are available, but that they are not complete, e.g. that we will not have data on the distribution of the audience of every vehicle across every conceivable population type and that we must have

some method for estimating these missing statistics if we are to form a working model of the mass media system. Professor Frederick Mosteller of Harvard University has suggested a parameter estimation iteration for generating full tables from such partial tables.

Let us look now at the kind of results or output which we might like from a simulation of a mass media system. What uses will we make of such a simulation and what kinds of manipulations shall we attempt with it? We should like to build a model or a black box which will answer the question of who gets exposed how frequently over what period of time via what channel to what messages in the mass media. At a later stage we may want at least implicitly to go beyond this to examine the effect of those exposures; at this time however we limit ourselves to the question of exposure. We may want, as with the present research, to do a content analysis of some mass media system to produce a scenario of real messages from which we may predict the rates of exposure, frequencies of exposure, and distribution of exposures in the population of that mass media system. Or having once established the validity of the model, we may want to undertake studies of hypothetical sets of messages in the media system to see what kind of exposures we would expect under different conditions. For example, we might wish to discover which strategy would be most effective in exposing those least

informed about some international issue to messages explaining the nature of that international issue. A model producing a record of exposures in the manner we have described above could be quite helpful in planning such a campaign, showing the most efficient way to invest resources to achieve the desired kinds of exposure.

How shall we build this kind of model? One approach would be to collect average exposure data for each relatively homogeneous subgroup of the population and then to aggregate these data over messages and subgroups to produce the desired exposure information. Let us consider which statistics would be required for some relatively homogeneous subgroup in the population of the media system. As a theme of messages appears we would like to know for each subgroup the frequency distribution of exposures at each time period of the simulation. Since data are usually available only by vehicles and since themes of messages will normally appear in several vehicles, in order to produce this frequency distribution of exposures, we would be forced to know for a particular vehicle the distribution of frequencies of exposure in each subgroup after the appearance of a set of messages in the vehicle and then aggregate over the vehicles in which the messages have occurred. Of course, data of this detail on the frequency distribution of exposure after n messages in a vehicle are never available. The closest approximations are the cumulation data

of vehicle (not message) exposures for the total population (not for subgroups). These data would be very hard to measure in principle since the messages in the various vehicles probably have widely varying attention-commanding characteristics, characteristics partly a function of the vehicle, partly a function of the display and location of the message in the vehicle, partly a function of the theme or content of the message, and all of these varying also with the characteristics of the audience.

Our solution to these complexities has been to approach the phenomenon of exposure as a sequential probability process. A given individual has a certain probability of exposure to a vehicle and a conditional probability of exposure to a message in the vehicle depending upon all the factors described above, given that he is first exposed to the vehicle. The net result is that a person's total probability of exposure to a given message is a product of two probabilities; his probability of exposure to the vehicle multiplied by his probability of exposure to the particular message given exposure to the vehicle carrying the message.

Even with this formulation we might, for each homogeneous subgroup of the population, derive a probability distribution for exposure to each vehicle for the member of that population subgroup. Then given a particular message with its conditional probability of exposure, we simply multiply the two numbers together in order to get the net

probability. However this method also has some problems; unless we derive some very complex joint probability distributions, we do not know the probability of a person's being exposed to one vehicle in relationship to his probability of being exposed to a second vehicle. Moreover, even if we could handle the data for each subgroup of the population in terms of a joint probability distribution of exposures over quite a lot of vehicles, we have absolutely no possibility of ever finding actual data, even by an estimation process, in the detail needed to describe these distributions for each population subgroup. For instance, if they exist at all, the duplication data between vehicles refer to the entire population and not for some subgroup of the population. Likewise, if cumulation data exist for a vehicle, they describe cumulation in the entire population; almost never do they describe cumulation in several subgroups of the population. Finally our routine for estimating missing parameters from known data requires a minimum amount of known data to be confident that the estimates are valid; the data on cumulation and duplication are so sparse that the validity of the estimation techniques seems quite doubtful. However if we could achieve duplication, cumulation, and average audience data for each subgroup, then we could have the simulation routines which now make probability assignments for the entire population call themselves and operate on the individual subgroup. In this case,

several hundred simulated individuals could represent members of that subgroup and the resulting statistics would be weighted by an appropriate factor to reflect the size of that population subgroup.

Because of the problems above we have compromised between designing a model which requires a complete description of the audience habits for each population subgroup and a model which operates at the level of the total population and ignores completely the need to differentiate between important subgroups. To describe the distribution of the population and the average audiences of the various communications vehicles over the population subgroups defined by the dimensions chosen by the researcher, we shall use the parameter estimation techniques on our initially incomplete data to provide these figures for each population subgroup. However we will use the considerably sparser duplication and cumulation data to generate vehicle exposure probabilities for the total population and then attempt to assign these probabilities to the population subgroups, consistent with the estimated subgroup parameters. Thus, since the data are so poor, we make no attempt to reproduce on a cell-by-cell basis the cumulation or duplication figures.

To summarize, a model of a communications system consists of:

1. an audience, each member of which has:

a. personal characteristics such as age, sex, occupation, place of residence

b. media habits, such as subscriptions

2. media:

each of which appears periodically in discrete issues, and each of which contains separate stories of items in each of which there are certain themes.

Among the most important quantitiative characteristics of a media system are:

1. the size of the population among whom the audience is found

2. the breakdown of that population into cells defined by the personal characteristics of the individuals

3. the audience of each vehicle (which is the circulation of that vehicle times the number of persons exposed to each exemplar of the vehicle)

4. the cumulation pattern of each vehicle

5. the duplication patterns between vehicles

The last two concepts, cumulation and duplication, need to be defined. Cumulation is the proportion of the population that has been exposed to a vehicle after n issues. For example, magazines tend to inform their advertisers as to the proportion of the American public who will have seen at least one out of one, one out of two, one out of three, or one out of four issues of the magazine. Duplication is the intersection between the set of persons exposed to one vehicle and the set of persons exposed to another. Thus American newspapers in multi-newspaper cities tend to

inform their advertisers as to what proportion of the population see their newspaper alone, see both newspapers, see only the other newspaper, or see neither. The extent of duplication is the extent of persons exposed to both.

The Simulation Model

We have discussed above the difficulties in producing a frequency distribution of exposures for each population type in such a way as to integrate all the known data about cumulation and duplication for several media vehicles. Our approach to this problem of integration of data at the subgroup level and data for the entire population is to use a probabilistic model in which each member of the hypothetical simulation population has a probability of exposure to each vehicle in the media system, these probabilities of exposure being chosen such that a) the average audience within the population subgroup or cell is correctly reproduced for each vehicle, b) the duplication of audiences across vehicles is correctly reproduced and, c) the cumulation in the entire audience for each vehicle is also correctly reproduced.

The simulation model is easily conceptualized in two separate parts; in the first part the hypothetical population with its probabilities of exposure to the communications vehicles in the mass media is constructed and stored on a disk or a tape in a computer. This hypothetical population consists of a correct proportion of individuals

of each population type with each person's probabilities of exposure to each of the vehicles in the mass media system. The second part of the simulation takes as input a scenario of messages organized by themes and time periods, which are then put into the mass media system to produce exposure of various kinds of individuals over time. However, this is not as simple as it may seem at first. Since the probabilities which have been stored for each person are only the likelihoods of exposure to the vehicle and are not likelihoods of exposure to the message contained in the vehicle, we must then estimate the conditional probabilities of exposure to the message given exposure to the vehicle. The parameters for calculating these probabilities of exposure to the messages given exposure to the vehicles are another input in the second part of the simulation.

The First Stage of the Simulation: Modeling the Vehicle Audiences

Now we will describe the sequence of operations used in constructing the hypothetical population with its probabilities of exposures to each of the vehicles in the media system.

Breaking the population into subgroups with homogeneous characteristics.--We start with a population of a certain size. The first step is to pick a set of attributes or dimensions which seem to explain much of the audience behavior in the mass media and then describe the breakdown

of the population in terms of these attributes. Since we also want to describe the vehicle audiences along these dimensions, the inputs in this stage are usually census-like tables by age, sex, literacy, region, education, etc. From these tables describing the breakdown of the population we construct a grand array of cells which is the intersection of all the classes of the attributes, giving the number of people in the population for every attribute combination. Often we arrive at a problem here: the tables in the various data sources which describe the population breakdowns rarely will include in one table all the dimensions which we wish to consider in the simulation. We usually find tables of three or four dimensions, at most. This is true for U.S. census tables and even more so for foreign census data. We are forced to construct a population based on these tables, taking all the information present, and deriving a reasonable population which might have given rise to the tables. For this, we use an iterative technique suggested by Professor Frederick Mosteller of Harvard which takes the known data and from it estimates the unknown parameters (cell sizes) for the table. The resulting values are those which contain all the information in the known tables, but no other information. They assume that other than the information in the known subtables, all interactions, or nonrandom effects, are zero in the population. Thus, even with incomplete tables describing the

population, we are able to make a reasonable guess about the distribution of the population across all of our relevant dimensions and create a table which includes all the data from the original tables.[7]

Distributing the audience over the population cells.--The next step is to perform exactly the same manipulations on the audience for each of the vehicles, distributing it over the population cells. For each cell describing a combination of attributes along all the dimensions of the population, we have the number of the population in that cell and the audience of each vehicle in that cell. If we divide the audience by the population in the cell, we then obtain for each vehicle the rating, or the proportion of the population of that cell which is in the average audience of the vehicle. At this stage, then, we have the population scattered into cells plus cell mean exposures, or ratings, for each vehicle for the cell. At this point we are ready to consider probabilities of exposure to the vehicles for each member of the population.

Modeling the cumulation process.--If we knew just the total population and the total audience for the vehicle, then the model with the least differentiation would assign

[7]A more complete description of the parameter estimation technique is provided in Chapter II.

every individual in the population a probability of exposure to the vehicle equal to the ratio of the vehicle audience to the total population. (Thus if the vehicle audience for a newspaper reached one-third of a population, then this model would assign to each member of the population a probability of exposure to that vehicle equal to one-third.) This model would indeed reproduce the average audience of the vehicle, but it would not usually reproduce the other data describing how the audience for the vehicle varies over the types of the population or its duplication with other vehicles or its cumulation over several issues. To use the average audience data known for each population type, we could take the cell mean audience as a mean probability of exposure for each population type. Therefore instead of one single probability for each individual, we would have as many probabilities as we have population types, assigning to each member of the population type the average probability for that population type. This is analogous to the first situation, except that now the population is differentiated into a large number of subgroups, each group having an average probability of exposure which generates average audiences equal to those average audiences which are either input or estimated for the population type for a given vehicle. However, this further differentiation would not necessarily reproduce the known or estimated cumulation over the entire population for a single vehicle,

nor would it necessarily produce a known duplication in the entire population for any pair of vehicles. Therefore we turn now to the cumulation data to calculate probabilities which will reproduce cumulation for each vehicle over the entire population.

The implications of cumulation for the simulation model are discussed more fully in Chapter IV and so we give only a brief outline at this point. If the average audience of a vehicle is one-third of the population, we can conceive of this happening in many different ways. Each member of the population could have a probability of one-third of being exposed to the vehicle every time the vehicle appears. In this case the composition of the audience varies randomly from issue to issue and the total number exposed at least once to the vehicle (the cumulative audience) grows with each succeeding issue. However there is another way to make the vehicle audience equal to one-third of the population: we could assign to one-third of the population a vehicle exposure probability of 1.0 and to the rest of the population a probability of 0.0. In this case, every time the vehicle appears one-third of the population would be exposed, but it would be the same one-third each time and there would be no cumulative growth in the number exposed. Obviously the cumulation we observe in the total audience is a function of the distribution of exposure probabilities in the population and by manipulating these probabilities

we can reproduce the known cumulation. This is the logic behind our use of cumulation data to develop probabilities.

For each vehicle, the population is divided into three groups. The first group is the distribution of very frequent consumers of the vehicle. These people might be the subscribers to a magazine or the habitual viewers of a television show. The second group is the distribution of those people who very infrequently are exposed to the vehicle. These people might be, for example, all the men in the population when the vehicle is a women's magazine, or members of the population who work at the time of a given television program, or those who do not own a set. The third distribution are those people with moderate probabilities of exposure to the vehicle, i.e., just the remainder of the population.

For each of these groups we must estimate the size, the average audience, and the two-period cumulative audience. Making the assumption that for each of these groups the distribution of exposure probabilities is of a known functional form--that of a mathematical function known as the beta function--then the average audience and the cumulation within each group is sufficient to calculate the parameters defining each of the beta functions, completely specifying the distribution of probabilities for the population for the given vehicle. These probabilities are then such that, for the entire population, they will reproduce the average audience and the two-period cumulation of the vehicle.

In calculating the distribution of the probabilities for a given vehicle, one probability for each member of the population, we have ignored our knowledge of the distribution of the audience of the vehicle across the population types and the resulting average probabilities of exposure within each population type. How shall we relate these many average probabilities of exposure to the set of probabilities generated from the cumulation data? The problem is to place individual probabilities from the cumulation distribution into the population cells, one probability for each person in the cell, such that these probabilities reproduce the known average probability for the cell. For cells that have a very high rating, or mean probability of exposure, it is obvious that the selection of probabilities to assign to these cells should come from the highest exposure group. In the same way, cells with very low mean exposure probabilities should have probabilities chosen from the low probability group. The actual assignment is made in the following manner: the mean exposure probability of the cell is compared with the means for each of the three distributions and a distribution is selected whose mean is closest to the mean of the cell. A probability is then drawn randomly from this distribution and assigned to the cell and eliminated from the distribution. This probability will then account for part of the audience expected in the cell. We can subtract this contribution from the

expected audience in the cell and calculate a new mean cell probability for the remaining probabilities to be selected. This new mean probability for the cell is then compared with the three exposure distribution means and the closest distribution is chosen as the distribution from which another random selection of a probability will be made. This procedure continues from cell to cell until all the probabilities from some distribution are exhausted and then the remaining closest distribution is used to draw the remaining probabilities. In this way, all the probabilities are assigned to cells, a probability for each person in the cell, although these probabilities are not yet connected to individual people. In Chapter VIII we show that this semi-random method for selecting probabilities does, in fact, fairly closely reproduce the required cell means as derived from the first part of the simulation. Of course, the approximation is better for larger cells, but is also true that larger cells are more important in the analysis of the data. Also, since the data will rarely be analyzed for each of 100, 200, or 300 cells, but generally for groups of cells, the resulting net assignment is even better.

Taking account of duplication statistics.--Finally we must manipulate these resulting probabilities in such a way so as to reproduce the empirical duplication of audiences between pairs of vehicles. We do this in the following manner; once the probabilities have been assigned to

cells of the population we can calculate the expected duplication if these probabilities for each vehicle were randomly assigned to individuals within the cell. Summing this duplicated audience over all the population types gives the duplication which would be produced by the simulation if the probabilities were randomly assigned within cells. If this duplication is significantly different from the empirical duplication then we must consider assigning the probabilities within the cells on a non-random basis. For example, if the empirical probability is higher than that which would be produced by the simulation using random within-cell assignment then we may want to increase the duplication within each cell by making people of high probability of exposure to vehicle A also have high probability of exposure to vehicle B. Recalling for each probability whether it comes from the high, middle, or low exposure group of its vehicle and also the average probability of each of these groups, we calculate the number of each possible pair of probability assignments (high-high, high-medium, high-low, etc.) for any two vehicles in order to most closely reproduce our empirically required duplication. These non-random assignments are made for pairs of linked vehicles (vehicles having a large non-random duplication) within groups of up to five vehicles at a time (since duplication may be related across several vehicles). A more detailed discussion of this algorithm follows in Chapter VI below.

These operations take all the known data, making estimates of unknown data from the known data where required, and integrate the data via the probabilities of exposure for each individual for every vehicle, to produce finally a set of simulated individuals, each of whom belongs to a certain population type as defined by the population dimensions and each of whom has a probability of exposure to each vehicle in the mass media. At this point, if we were to run a scenario of simulated messages, we could calculate expected exposures to each of the vehicles, the variance of this number, the duplication, etc. The next step is to generate probabilities of exposure to the messages given exposure to the vehicle. This process takes place in the second stage of the simulation.

The Second Stage of the Simulation: The Processing of Scenario Messages and Reporting of Exposure Statistics

Representing the message exposure probabilities.--In the second stage of the simulation, we calculate conditional probabilities of exposure to the message given exposure to the vehicle in order that our exposure statistics represent exposure to messages and themes of messages rather than exposures to vehicles carrying the messages. Conceivably, these conditional probabilities of exposure to a message can depend upon all of the various factors having to do with exposure in the simulation. They may be

a function of the population type of the individual, of the content or theme of the message, of the communication vehicle in which the message is carried, of the time period, of the history of previous exposure of the individual, etc. This could mean then, that for every message in the scenario, we have as many different message exposure probabilities as there are population types, which in the present simulation would be 144 message exposure probabilities for each message. This number of message exposure probabilities for each message is impossible to handle on the present computer.[8] However, there is a more important reason why we do not admit all these possible message exposure probabilities: we simply do not have the data available to specify for every population type a different message exposure probability. The knowledge we have of conditional exposure probabilities is very meager indeed, e.g., we can differentiate average exposure levels for broad categories of themes such as international news in general but not between different kinds of international news. For the broad category of

---

[8]Note also that we are talking only of the average probability of exposure in each population type; we are not taking into account cumulative conditional message exposure probabilities which would give a distribution of probabilities of exposure within each type. If we actually developed these distributions for each message, we would generate a number of probabilities equal to the size of the population, as we did in the first part of the simulation. In this case, however, instead of a number of such distributions equal to the number of vehicles, we could have hundreds of such distributions, one for each message.

international news we find some indication of exposure differences between people arranged on a single dimension, i.e., sex, education, age, etc., but not all the possible combinations. Also, the available data are all derived from newspaper readership, but do not distinguish between newspapers, nor do they distinguish any other media vehicles.

Given the kind of data available and the limited capacity for processing and storage in the present model, we have compromised in the amount of detail of the message exposure probabilities and used the following model; first, we assume that there is some intrinsic level of attention for each of the themes in the scenario. This attention level is an average message exposure probability over all possible formats, media vehicles, time periods, population types, etc., and the researcher is required to specify this average level of exposure for a theme. We know from research that we can make such statements as, "Twenty percent of the people who read newspapers regularly, read international news in the newspapers," and this kind of figure is what we mean by the average exposure to a theme, given exposure to the vehicle.

Now consider any dimension which we have reason to believe influences exposure to the message given exposure to the vehicle, e.g., education. Each such dimension has a number of classes or levels. The second assumption is

that for a given vehicle, the ratio for any two levels of the average probabilities of exposure to the message given exposure to the vehicle is constant for all messages of the theme. For example, we might want to assume for any message relating to a theme concerning the United Nations, that highly educated people are twice as likely to be exposed to the message, given that they are exposed to the vehicle, as are poorly educated people who are exposed to the vehicle. This assumption says nothing about the absolute level of exposure to the message; it is assumed that the absolute level of exposure to the message is governed by the general interest inherent in the theme and by the format of the message in the vehicle. Therefore, for each theme we must estimate a general level of attention which is particular to the theme (called PORTN in the simulation) and ratios of exposure probabilities for people in adjacent levels of the relevant dimensions.

We use these ratios as input because the researcher can probably more accurately estimate a ratio of two exposure probabilities than each value separately, and also because it is not possible for the researcher freely to specify both the average conditional audience for a theme and also a set of probabilities of exposure over a population dimension: these figures are related and must be compatible. For example, if for men the conditional probability of exposure to the message, given exposure to the

vehicle is 0.30, and the corresponding probability for women is 0.20, then the average probability of exposure or the average proportion exposed to the message, given exposure to the vehicle, lies somewhere between 0.20 and 0.30 depending upon the number of men and women in the vehicle audience. If the researcher has specified that for this particular theme the average probability of exposure for the vehicle audience is 0.27, then he is not simultaneously free to choose both the conditional probability of exposure for women and that for men, since the weighted average of these probabilities must equal 0.27. The researcher usually does not know these conditional exposure probabilities to the degree of accuracy with which he knows the overall level of exposure for the theme, and he doesn't need to be bothered with knowing the breakdown of the vehicle audiences for every vehicle.

Knowing the average proportion of the vehicle audience exposed to a given theme, the ratios of exposure probabilities for those population dimensions chosen by the researcher, and the distribution of the vehicle audience along those dimensions, we can calculate for each vehicle the theme's message audience distributed across each of the dimensions. The distribution across the population subgroups defined by the set of dimensions can be calculated either by cross-products or using the Mosteller iteration: if we use the iteration we can distribute the theme's

message audience biased somewhat according to either the vehicle audience or the population distribution in the subgroups.[9] Finally, the ratio of the message audience to the vehicle audience in a subgroup is, for that subgroup, the average conditional message exposure probability, for that theme and vehicle.

How then do we account for differences in exposure from message to message within a theme? Each message in the theme is described by a format factor, a positive number by which we multiply the value of the average audience of the theme in order to get the audience for the particular message. Thus, the average proportion of the vehicle audience in the message audience for the theme (PORTN) is the proportion for a message with a format factor of 1.0. Messages which are quite prominantly displayed with pictures, headlines, etc., have format factors greater than 1.0 and messages which are buried inconspicuously somewhere in the newspaper have format factors less than 1.0. If we were to do a study to assign format factors to messages, we would try to develop from the formating for each message a multiplicative factor by which we could characterize the message and predict the exposure to it.

Of course, this model is a simplification. For instance, we have reason to believe that the exposure to

---

[9] This technique is discussed in Chapter II.

general international affairs news is about 1.8 times as high among men given exposure to the vehicle as it is among women; therefore, for international affairs themes we use this probability ratio for every message. However, it might very well be that the ratio changes, depending upon the format of the message, for example whether the message is on the first page of the newspaper or in the women's pages. We do not take account of this possible interaction between format and probability ratios except insofar as the averages resulting from the numbers of messages on the front page and the numbers of messages in the women's pages give an average level of exposure to the theme. Nevertheless we will consider this model sufficient since we actually have only poor data with which to estimate the exposure probabilities for the messages. Since, in the present simulation, the probabilities are themselves derived from a non-random sample by a regression procedure which explains only 40 to 50 percent of the variance, we feel that the model is sufficiently complex for the level of validity of the data and that the errors introduced will generally be small since the mean exposure will be fairly accurately represented.

Thus, the data required for the last stage of the simulation consists of a scenario of messages organized by theme and into time periods, each message of which carries the identification number of a communication vehicle in

which it appears and a format factor describing the formating of the message in the vehicle in terms of the model explained just above. In addition, for each theme we estimate the average proportion of the vehicle audience exposed to the theme and the ratios of conditional exposure probabilities for the media vehicles along any set of the dimensions by which population types are defined until the capacity of the machine is used. Finally, the exposure portion of the simulation is run, processing each message in turn, multiplying the format factor, the conditional probability of exposure appropriate to the population type of each individual, and the vehicle exposure probability, to generate for each person-message combination a net probability of exposure.

Triggering of exposures and the notion of trigger themes.--It seems likely that there are certain single messages or certain themes which, upon being exposed to such messages or themes, increase one's probability of exposure to other messages and themes. A single exposure to the news of the assassination of a president would be enough to significantly change the immediate behavior of anyone so exposed, causing him to attend to much a greater degree to information in the mass media. In other words, both the vehicle and the message exposure probabilities for most individuals would increase. To account for this

effect, the simulation has a device which allows the researcher to change (increase or decrease) either the vehicle exposure or the message exposure probability or both, based either upon previous exposure to some theme or upon the time period. We allow the researcher to specify as many as three themes, called "trigger" themes, exposure to which (a given minimum number of times) changes the probability of exposure to other themes in subsequent time periods.

In order to use this facility the researcher must specify which of the first three themes are trigger themes and a number of exposures to a trigger theme which will trigger increased exposure to other themes. This triggering may be based on past exposure to any of the trigger themes and/or the current theme and/or the time period. The amount of the change in each of the probabilities may be a function of the media type of the message, of the time period, or of the population type of the person being exposed. The direction and the magnitude of the change of the probability is specified by the researcher and includes a parabolic ceiling and/or floor effect which prevents probabilities from becoming greater than 1.0 or less than 0.0.

The modification of probabilities based on time periods is intended to allow the researcher to account for exogenous variables such as an initial exposure to a message via word-of-mouth communication. The change in

probabilities (the "triggering" effect) can be reversed slowly or abruptly as desired by the researcher. In fact it is possible to impose up to three cycles of triggering in the simulation run. Since the themes of the present simulation concern common day-to-day news events we have not used the triggering effects in our simulation runs.

Processing the messages and outputting audience exposure.--At the output stage of the simulation, each individual's probability of exposure to each message is calculated. As originally conceived, this probability would be used in a Monte Carlo routine to determine whether an exposure had taken place for the given person and message, resulting in survey-like records of exposure. As with survey data, these records would then be aggregated in many different kinds of tables. However, with this method of producing exposure tables, it is necessary to perform (at a minimum) several runs of the simulation and average the results to get some indication of the expected outcomes and the range of possible outcomes. Since a typical simulation (involving up to 3,000 persons and 1,000 or so messages over 10 time periods) requires several hours of computer time, it is quite costly to use the Monte Carlo solution to the simulation.

Fortunately, there are at least two other techniques for generating or (more accurately) describing simulation outcomes. These techniques are doubly advantageous in that

not only do they require only one run of the simulation, but also they produce more information about the likely outcomes of the simulation. The first of these techniques calculates, for each of the possible outcomes of the simulation, the probability of that particular outcome. Thus, we might have for each individual the probability of his being exposed exactly once, twice, three times, four times, etc. to the third theme of the simulation by the fourth time period. This kind of information is especially useful in testing the event validity of the simulation, i.e., in testing the simulation against real data. For example, if we had data demonstrating that a particular member of the NORC survey panel was exposed three times to the second theme by the fourth time period and the simulation predicted a relatively high probability of exactly three exposures for the person, then we would increase our confidence in the validity of the simulation.

Most simulations are programmed to report only a single outcome even though probability estimations are explicit or implicit in the model, e.g. an event is reported as surely happening when a threshold of some index or scale is passed. This kind of reporting makes validation somewhat difficult since it does not include reports on the likelihood of other outcome states of the simulation. In terms of the example above, if of a large number of possible exposures, the simulation gave our person a highest

probability of two exposures and a second highest probability of three exposures, we would lose much information if it only reported two exposures: we might, if given full information, consider the simulation to have significant validity. Thus, knowing the probabilities associated with each of the outcome states is essential for assessing the degree of validity of the simulation.[10]

Describing the outcome states of the simulation and their associated probabilities is the most complete description possible of the simulation outcomes. However often it requires much computation to arrive at the probabilities for each of the states. In the present simulation, the computation becomes unmanageable because of the essentially aggregate nature of the outcomes. It makes no sense to identify particular individuals in the simulation population with particular individuals in the NORC sample panel with which the simulation is to be validated, because the whole modeling of the simulation has taken place in terms of subgroups or cells of people, all of whom are taken as

---

[10]This approach to validation of complex models is described in some detail-including such notions as the descriptive power as contrasted with the predictive power of a model-in three papers by Joseph F. Hanna. The first of these papers is entitled "A New Approach to the Formulation and Testing of Learning Models" published in Synthese, Volume 16, 1966, pages 344-380. The second is Some Information Measures for Testing Stochastic Models (Michigan State University: mimeo, 1967), and the third is Information-Theoretic Techniques for Evaluating Simulation Models (Michigan State University: mimeo, 1969).

equivalent within each cell. Thus, the simulation data which can be meaningfully compared with the real world panel are aggregate statistics for the cells of the simulation. An example of such a statistic might be the probability that between 25 to 35 percent of the fifth cell of the simulation population is exposed to exactly 0, 1, 2, 3, 4, or 5, etc. messages for a particular theme and time period. This kind of statistic is very hard to compute exactly because it involves the probability distribution for the aggregate of people in the cell over all possible outcomes for a set of messages.

In order to avoid this difficulty, instead of calculating the values of the probability distribution for a cell, we calculate an expected outcome for the cell, i.e. we take the expectation of the probability distribution which is so difficult to calculate.[11] In this framework the previous statistic would become the expected number of persons exposed exactly once in the fifth cell of the population. However, at this point, we must acknowledge a difficulty in the use of this kind of a statistic in the

---

[11]It often happens that the calculation of a probability distribution is very difficult and time consuming, or requires an inordinate amount of space in the computer, but that the calculation of certain parameters such as the expected value is quite feasible and straightforward. For a simple illustration of this situation, see the short article by Ronald A. Howard entitled "Stochastic Process Models of Consumer Behavior" in the <u>Journal of Advertising Research</u>, Volume 4, 1966, pages 35-42.

present simulation: our expected values are not quite actually expected values. In the first stage of the simulation a random process was used to distribute probabilities within and across cells. Thus, the expected values which are generated in the second stage of the simulation as output data are conditioned upon the assignment of probabilities within and across cells which occurred in the first stage of the simulation. There are constraints upon the assignment of probabilities within cells, namely that the probabilities must come close to reproducing the cell means and that the probabilities are chosen from a prescribed distribution generated by the cumulation data. Because of these constraints and because the expected values are almost always reported for larger population subgroups (i.e. combinations of the 144 basic cells), we find (in Chapter VIII) that the conditional expected values are rather close to the true expected values given all possible assignments within the assignment process.

Ideally, with each expected value we would report another statistic, the variance, which would help to define the probabilities of the outcome states of the simulation.[12] If we know the expected number of exposures in a subgroup

---

[12]Unfortunately in programming the current version of the simulation we did not realize the importance of the variance here, and so we did not report it. Also however, it would have been quite difficult to find storage space for twice the present number of statistics.

and the variance in the number of exposures, we have a better idea about the range of likely outcomes about the expected outcome. It may also be true that most of the distributions for which we report expected values approximate normal distributions, in which case the knowledge of the expected value and the variance would be sufficient to describe completely the probability distribution over all of the possible outcomes of the simulation for these statistics.[13]

The Exposure Statistics

What kinds of statistics should be output from the simulation? Because the processes of the simulation operate at the level of the population subgroup, we have chosen to report summary statistics for the subgroup, i.e. the expected number exposed and the average expected number of exposures. These statistics are calculated for each time period and also cumulatively. We do not report the distribution of exposure within a subgroup, although we now feel

---

[13]Since the messages have message exposure probabilities which differ from message to message, the probabilities are not constant for each trial and the probability distribution for the number of exposures is very complex. Nonetheless, since there may be several hundred or more messages in a single theme, it seems plausible that the probability distributions might very well approximate normal distributions. (The Central Limit Theorem may be relevant here.) The matter of the approximation to the normal might be an area of fruitful research in the future because it would mean that by including the variances as output in the statistics, we could quite significantly improve the description of the output of the simulation.

that this is an important statistic. Each of these statistics, however, could possibly be reported by a bewildering array of other attributes, e.g. by message, theme, time period, vehicle, audience type, cumulation distribution type, or by combinations of themes (to get duplicated audiences). In addition, the variance of each statistic could be reported.

A rough calculation, then, gives a possibility of hundreds of millions of different data items which might be of interest to the researcher. Of course, this is an impossible number of data items to deal with and, in fact, we must make some simplifying assumptions in order to be able to offer some reasonable possibilities to the researcher and still keep the program manageable. One of the simplifying assumptions is that messages, although input as individual messages, are organized by themes, and in the output phase an individual message is not distinguished unless a particular theme has only one message in it. Therefore, instead of 300 or 400 different messages to be considered as output, we actually consider some much smaller number, say 15 or 20 themes. We have noted above that we do not report the exposure frequency distributions or any variances. Also, most statistics are not output by 64 vehicles, although basic exposure by vehicle is given. We have defined media types, of which there may be at most six, which group the vehicles together: therefore, the

64 possible vehicles may be grouped into any one of six media types and most statistics are output by the six media types.

This is the extent of the simplifying of the output data. It nevertheless leaves a huge amount of possibilities for output, and the computer could obviously generate such a large number of statistics that it would inundate the researcher. Therefore, the researcher, even at this stage, is asked by the simulation to specify the tables of interest to him. The most important choice is the population subgroups for tables. It seems quite unlikely that with 144 population types, as in the present simulation, a table printed out with 144 values could be comprehended by the researcher. Therefore, these tables can be collapsed, and one may ask for any combination of tables with any order of dimensions, e.g. tables of two dimensions, three dimensions, four dimensions, and so on, as desired, rather than looking at the table of highest possible dimensionality. Below we list the resulting output statistics for the simulation.

A Summary of the Simulation Output Statistics

1. The number of exposure events by media vehicle.
2. The cumulative number of exposure events by media vehicle.
3. The number of exposure events via the media type.
4. The cumulative number of exposure events via the media type.

5. The number and percentage of each population type exposed.

6. The cumulative number and percentage of each population type exposed.

7. The number and percentage of each population type exposed via media type.

8. The cumulative number and percentage of each population type exposed via media type.

9. The total and average cumulative number of exposure events for each population type.

10. The total and average cumulative number of exposure events for each population type via media type.

11. The cumulative number and percentage of each population type in the audience of both the current theme and a "trigger" theme.

A Summary Outline of the Mass Media Simulation

One common way of approaching a piece of research in social science is to ask what are the independent and dependent variables treated in the study. In addition one should ask what kind of assumptions or models allow us to relate the values of the independent and dependent variables. The table of output statistics above may be considered a description of the dependent variables in the simulation; we present here a brief summary of the independent variables and the assumptions which relate them to the dependent variables.

A Summary of the Independent Variables

1. The (sometimes partially) known distribution of the population across the dimensions chosen by the researcher.

2. The (sometimes partially) known distributions of the media vehicle audiences across these dimensions.

3. For each vehicle, the breakdown of the population into high, medium, and low vehicle exposure groups and the average and two-period cumulative audience for each group.

4. The known duplicated audience for any pair of vehicles.

5. A set of messages, organized by theme and time period, with a media vehicle identification and a format factor for each message.

6. For each theme, the average proportion of the vehicle audience exposed to the theme.

7. For each group of vehicles, for each theme, for some subset of the population dimensions, the ratios of conditional message exposure probabilities of adjacent classes along each dimension.

Assumptions in the Simulation Model

We attempt to enumerate here the most important assumptions relating either to the simulation model, or relating to the data of this particular simulation or model instance, or combinations of the two of these.

1. We assume that the combination of our incomplete data and the Mosteller parameter estimation iteration which produces maximum likihood estimates of population and audience values in the cells of the simulation is sufficiently accurate for estimating unknown parameters. This is equivalent to assuming that unknown interactions in the data can be neglected, either because they are very small, or because the communication vehicle or population subgroup involved is not of major importance for the particular scenario of messages whose flow we are attempting to simulate.

2. We assume that we can model the flow of messages to each individual as a Bernoulli process in which a) the probability of exposure to one message by an individual is independent of the probability of

exposure to another message by the individual, b) the probabilities are independent from individual to individual, and c) these probabilities do not change over time or at least during the duration of the time simulated by the scenario of messages.

3. We assume that the three beta function model of the distribution of these probabilities for each communications vehicle adequately describes the individual vehicle exposure probabilities for the population.

4. We assume that the quasi-random process of assigning probabilities to cells from the cumulation distribution makes a realistic integration of the population cumulation data and the cell audience data, i.e. that the process sufficiently well reproduces the subgroup average and cumulative audiences.

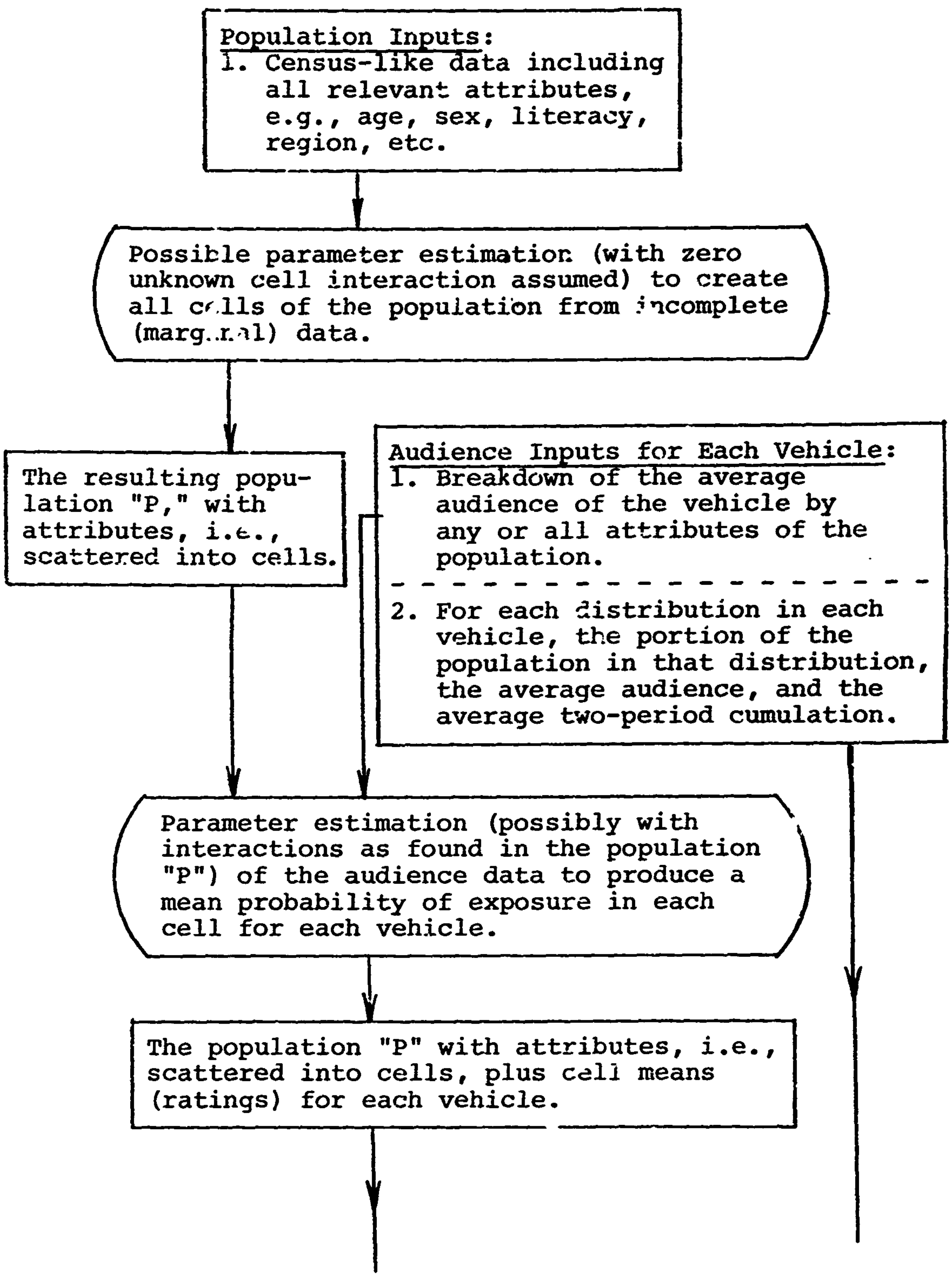
5. We assume that the vehicle audience duplications not accounted for by the cell mean exposure probabilities can be reproduced by the non-random within-cell assignment.

6. We assume that the model of theme-based conditional message audiences with individual message audiences produced (multiplicatively) by format factors and a single average conditional message exposure probability for a population subgroup adequately fits a wide variety of empirical exposure data.

Finally, we present flow diagrams of the two stages of the simulation. In the first stage, the simulation may be thought of as a very efficient data disaggregator-integrator, or a consistency machine. Such diverse data as population tables, audience and cumulation subtables and marginals for each of several media vehicles, plus audience duplication figures between pairs of media vehicles are transformed into probabilities of exposure events assigned to members of a model population on the basis of the several assumptions described above. The flow diagram of Figure I-1

shows the transformation of the data. The most important theoretical assumptions are represented by ovals in the flow diagram.

The second stage of the simulation is the message exposure and reporting stage. Messages are processed time period by time period and theme by theme and the growth of exposure is reported.


Population Inputs:
1. Census-like data including all relevant attributes, e.g., age, sex, literacy, region, etc.
Possible parameter estimation (with zero unknown cell interaction assumed) to create all cells of the population from incomplete (marginal) data.
The resulting population "P," with attributes, i.e., scattered into cells.
Audience Inputs for Each Vehicle:
1. Breakdown of the average audience of the vehicle by any or all attributes of the population.
2. For each distribution in each vehicle, the portion of the population in that distribution, the average audience, and the average two-period cumulation.
Parameter estimation (possibly with interactions as found in the population "P") of the audience data to produce a mean probability of exposure in each cell for each vehicle.
The population "P" with attributes, i.e., scattered into cells, plus cell means (ratings) for each vehicle.

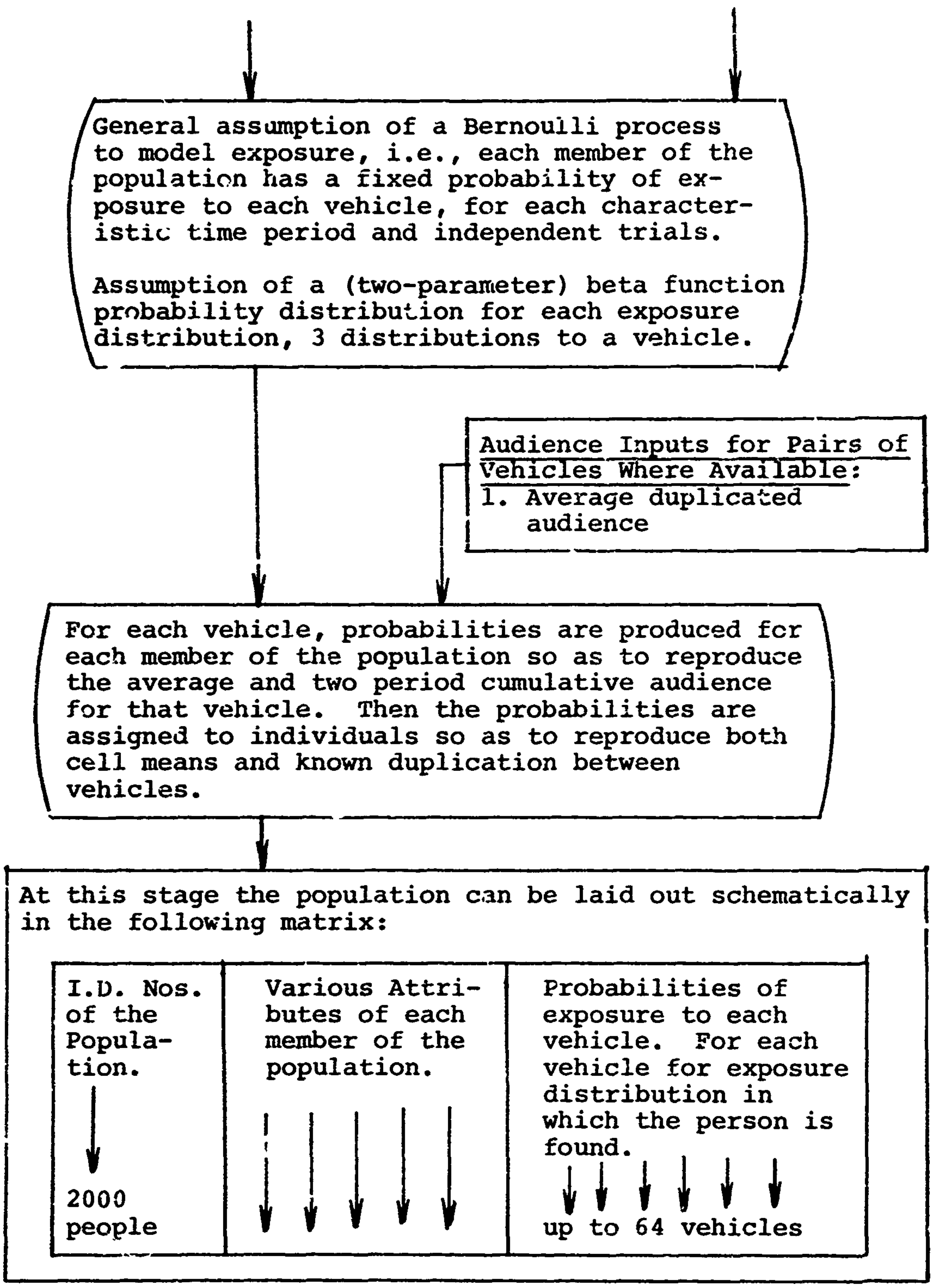


Figure I-1. The First Stage of the Simulation: Data Disaggretion and Integration.

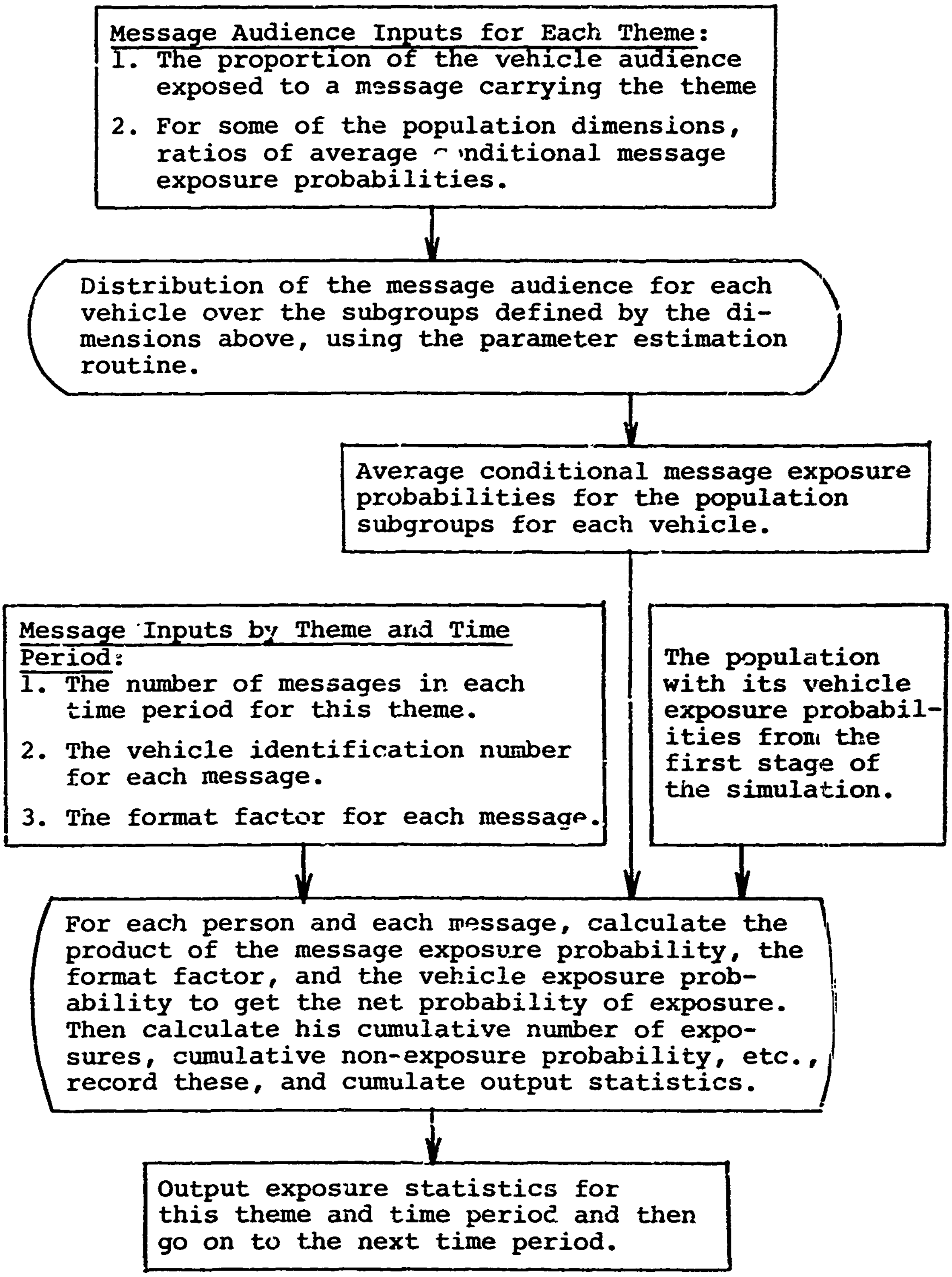


Figure I-2. The Second Stage of the Simulation: Processing Messages and Reporting Exposures.

CHAPTER II

THE MOSTELLER PARAMETER ESTIMATION TECHNIQUE

In the first chapter, we discussed several situations in which we had to estimate cell entries in multidimensional tables from subtables of lower dimensionality, e.g., in forming the population and in distributing the vehicle audiences and the message audiences. There are a variety of such situations in general in which the social scientist might use such a parameter estimation routine.

Sometimes the data available describing the population is from several different sources, each one of which has some but not all of the dimensions of interest to the researcher. This is a very common occurrence. For instance, in the case of the Soviet population, we have census data by sex, age, residence, but not by party affiliation. However, there are data available from other sources on the breakdown of party members by sex, age, education, literacy, etc., and we might wish to combine these data to construct a model of the Soviet population.[1]

---

[1]Herbert J. Selesnick is presently simulating Soviet exposure to news of two international crises and has encountered this problem in the process of building a simulation population.

The technique which we shall describe below allows one to join these subtables together to produce a population which reflects all the interactions in the known data and no other interactions. Of course, this is only possible provided that the subtables are themselves consistent. We shall discuss at some length this problem of consistency or inconsistency of data from several sources.

Another kind of problem for which one might use the parameter estimation routine is the problem of sparse data in contingency tables, where the cell sizes are so small that, taken individually, we can have relatively little confidence that they are precise enough estimates of the underlying population. One way to handle this problem is to collapse the contingency table, either collapsing out entire dimensions or collapsing categories within dimensions, to provide more stable estimates of the resulting subtable values. If, however, for purposes of simulation or calculation of rates or whatever, we require the table in its highest dimensionality, this procedure will not do. A solution to this problem is to perform the collapsing of certain dimensions of the contingency table in order to get the more precise estimates of the cell values. Then we assume no interaction at the higher dimensionality level and, on the basis of this assumption, generate expected values from the lesser dimensionality tables in which, presumably, we can have more confidence. This procedure

produces smoothed contingency tables based on the assumption of no interaction at a level of dimensionally higher than that represented in the tables which were produced by collapsing out some subset of the original dimensions.

Mosteller characterizes this approach as

> . . . borrow[ing] strength for the cell estimation from the margins by giving up the contributions of the higher order effects . . . This is not to deny the possible existence of such [higher-order] interactions. Rather the thought is that it may be profitable for cell estimation purposes to ignore them. A simple analogy arises in ordinary regression where we often fit a straight line or a parabola when we know very well that the true state of affairs is more complicated, and sometimes even when we have a clear picture of how the true state of affairs should be represented. We do this because our estimation may be better when using a simple model in a limited range, even though the model is wrong, than when fitting the correct model.[2]

In the present simulation, we use the parameter estimation technique in both of the ways described above. In calculating the population and daily newspaper audience distributions, we use data from the NORC sample which was too small (N=745) to provide stable cell estimates for the 144 population cells. Therefore, we collapse the five-dimensional table to the ten three-dimensional subtables from which we make "smoother" estimates of the cell entries in the five-dimensional table. For the radio audiences we simply use our estimated subtable values directly in the iteration.

---

[2]Frederick Mosteller, "Association and Estimation in Contingency Tables," Journal of the American Statistical Association, 63 (1968), p. 19.

Let us describe the theoretical foundations for this iterative technique for estimating cell values. First, we introduce some notation to deal with the problem. In Fig. II-1, we have drawn a 2x2x2, three-dimensional contingency table as a cube with three faces. The values in the interior cells of the cube are labeled with x_{ijk}. Here, i represents the value of the level of the first dimension, j is the level of the second dimension, and k is the level of the third dimension. If we sum the x_{ijk} over all the levels of any one dimension, the resulting values will be represented with a plus sign (+) at the place in the subscripts corresponding to the dimension which has been summed out. Summing out a dimension leaves a new set of values and if it is only one dimension that is summed out, the new set of values are just the values which we would put on one of the faces of the cube. Thus, the cube below has three faces, x_{ij+}, x_{i+k}, and x_{+jk}.

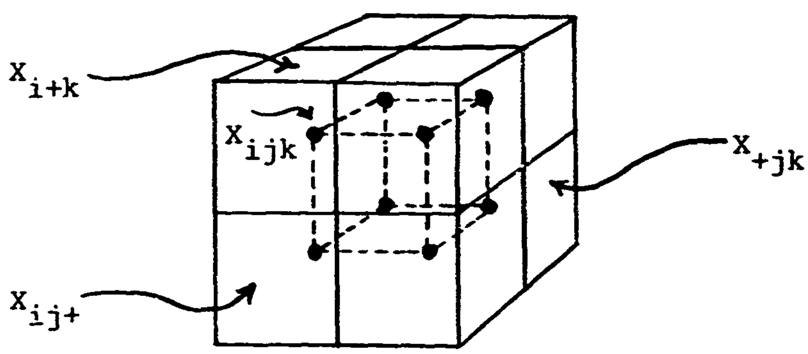


Fig. II-1. A Representation of a 2x2x2 Table as a Cube.

If we are willing to assume that there are no third order interactions, i.e., that the only interactions in the table are represented by the values on the faces of the cube, then Birch[3] has shown that a) the marginal totals (the totals on the faces) are maximum likelihood estimates of their expectations, b) there is a unique point for which the likelihood function is a maximum, and c) the maximum likelihood estimates for the interior cell values are determined uniquely by the appropriate marginal totals (by the values on the faces) and these marginal totals are preserved. This means that if we know only the faces of the cube and are willing to assume that there is no third order interaction, then the values of those faces under a variety of sampling conditions are the sufficient statistics for the likelihood function and from the face values we can generate a unique most likely set of interior cell values. These interior cell values will sum appropriately to preserve the known cell values.

When can we make the assumption that there are no interactions other than the known interactions? In general, this depends upon the situation for which we are using the estimation technique; however, if we have a choice about how many dimensions to include in the faces, Birch

---

[3] M. M. Birch, "Maximum likelihood in three-way contingency tables," Journal of the Royal Statistical Society, Series B, 25 (1963), 220-33.

postulates a hierarchy principle. The principle implies that if a lower order interaction is zero, then all higher order interactions involving the same combination of variables are also zero.[4] One can give examples of strange contingency tables where this hierarchy principle is not true, but it seems a reasonable presumption in the face of no other knowledge. Of course, if we are in the position of the researcher who has gathered together all the subtables he can about a population and is not smoothing data, but trying to generate absolutely unknown data, we will usually be forced to assume that there are no other interactions. We must be satisfied with a population which reflects all the known data and nothing more.

The iteration which calculates interior cell values from the faces proceeds in the following manner: first we initialize the interior values, usually by giving them all an initial value of 1.0, i.e., no interaction. Then, beginning with one of the cells of the known subtables, e.g.,

---

[4]Mosteller's student Bishop has investigated the choice of lower dimensional tables to estimate values in higher dimensional subtables for the Halothane study. See Yvonne M. M. Bishop, Multidimensional Contingency Tables: Cell Estimates, Ph.D. Thesis, Department of Statistics, Harvard University, Cambridge, Massachusetts, February, 1967; and also Chapter 4 of J. P. Bunker, W. H. Forrest, Jr., Frederick Mosteller, and L. D. Vandam (editors), The National Halothane Study. A study of the possible association between halothane anesthesia and postoperative hepatic necrosis. Report of the Subcommittee on the National Halothane Study of the Committee on Anesthesia, National Academy of Sciences--National Research Council. In press, will be available from Division of Medical Sciences, National Research Council, 2101 Constitution Avenue, Washington, D.C. 20418.

one of the cells on a face of the cube, we sum the appropriate interior values and compare the sum with the required cell value in the subtable. If the sum differs from the required value, we multiply each of the interior cell values included in the sum by the ratio of the required value to the sum, thereby normalizing each of these included values so that they do sum correctly. In terms of our cube above, if x'_{ij+} is the sum of the initial interior values, then

$$x'_{ij+} = x_{ij1} + x_{ij2}.$$

We would then create new interior values x'_{ijk} by multiplying

$$x'_{ijk} = x_{ijk} \cdot [x_{ij+} / x'_{ij+}].$$

We proceed in this fashion until the sums fo all the values in the subtable (on the face) sum correctly. Then we go on to the next subtable or face. Of course, as we process the second subtable, we will change the interior values calculated from the first subtable, and so the first sums will no longer agree with the first subtable values. However, as we repeat this process through the subtables several times, the interior values will quite quickly converge to values which sum properly for each subtable, <u>if the subtable values are consistent</u>. To the extent that the subtable

values are inconsistent, the interior values can never satisfy all the subtable values.[5]

The proof of convergence of the values (with consistent subtables) and an interesting interpretation of the final values is provided by some recent work in information theory. Each of the subtables used in the iteration contains a certain amount of information (in the technical sense of the word) about the population. Brown showed that the iteration procedure converges with each step to the ". . . minimum information (i.e., maximum entropy) extension . . ." of these subtables, i.e., the only information (non-randomness) in the final distribution is that contained in the subtables.[6]

The minimum information values are achieved when the initialization of the iteration contains no information, e.g., when all 1.0's are used as an initial starting place. However, it is possible to start the iteration with other values which do contain information, and thereby achieve final cell values which combine both the information in the

---

[5]The iteration was first proposed by Deming and Stephan in 1940. See W. Edwards Deming, and Frederick F. Stephan, "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known," Annals of Mathematical Statistics, 11 (1940), 427-44; and W. E. Deming and F. F. Stephan, "The sampling procedure of the 1940 population census," Journal of the American Statistical Association, 35 (1940), 615-30.

[6]David T. Brown, "A Note on Approximations to Discrete Probability Distributions," Information and Control, 2 (1959), p. 386.

subtables and that in the initialization. As an example, sometimes we have data about a population which is helpful in describing interesting interactions, but which fail to agree with other, more substantial data which itself does not give evidence about these interactions. We would like to combine both sets of data. In later simulations of audience exposure in Communist China, we shall make use of data from several different sources. One source of data is census reports which describe the population breakdowns along several dimensions for the population of China. We are willing to make the assumption that these census data are as accurate as any data that we are likely to have and we will accept these data at face value. These data, however, are never found in tables of dimensionality higher than two or three and do not represent the interactions between sex and literacy, sex and education, age and education, etc. We do have other data from interviews with refugees in Hong Kong in which the interactions between the various population dimensions are measured; however, the marginal values of this sample are not representative of the marginal values for the population of China as a whole. We would like to combine the interactions of the sample with the known marginals given in the census data for China.

We can use the iteration to produce this combination by initializing the interior cells with the data describing the interactions rather than with all 1.0's

(which imply no interaction). Then we use the known marginal data as subtables and proceed with the iteration. In this case the final values are not the minimum information extensions from the subtables, but combine the information in the two kinds of data. Mosteller offers an interpretation of this process:

> . . . it is reasonable to think of the association between [two variables] as independent of the relative numbers [in the classes of each variable]. Why should tripling the number of A's and halving the number of not-A's in each category . . . have any effect upon the basic association in the table? One reasonable answer is that it should not. We might think of a contingency table as having a basic nucleus which describes its association and think of all tables formed by multiplying elements in rows and columns by positive numbers as forming an equivalence class - a class of tables with the same degree of association. . . . an index of association [for a 2x2 table] which is invariant under these row and column multiplications is the cross-product ratio
>
> $$\alpha = \frac{ad}{bc}$$ [7]

In Mosteller's terms, the interviews of Chinese refugees capture (we assume) the basic nucleus of association in the population.[8] We only need calculate the

---

[7]Mosteller, "Association," p. 4. The italics are Mosteller's.

[8]We must be cautious about this assumption. Mosteller notes that, ". . . making the assumption that when a subgroup is formed from a parent population the multiplicative invariance is preserved does not make it true (as a vehicle for getting an estimate in ignorance, it has the same status as the use of linear regression in the absence of knowledge of the shape of the function)." Mosteller, "Association," p. 10.

equivalent table having the proper marginals to model this population.[9]

In the present simulation we will use this technique for calculating the distribution of the audience of the Sunday *Enquirer* over the population. We have data on the distribution of the daily *Enquirer*, but none for the Sunday *Enquirer*; however, we know that the audience of the latter is about sixty-four percent greater than that of the former. Therefore, we have increased each of the subtable values for the daily *Enquirer* by this value and used these values in the smoothing operation for the Sunday *Enquirer*. This process may, however, produce audiences in some cells greater than the population. We therefore initialize the iteration with the population values to combine the information about the distribution of the population with the audience estimates.

The Consistency of Subtables

In the process of using the parameter estimation technique at various stages in the simulation, we have discovered some very intriguing facts about the consistency

---

[9] We often encounter the need to synthesize data in this manner when we are doing secondary analysis of survey data. Ithiel de Sola Pool shows several examples in his "Use of Available Sample Surveys in Comparative Research," *Encyclopedia of Sociology* (1963), pp. 16-35. An interesting discussion is also presented by Leslie Roos in *Intersurvey Comparison - A Progress Report*, mimeo (Massachusetts Institute of Technology, 1967).

or inconsistency properties of subtables. The iteration must converge if the tables used in the iteration are consistent. But, how can we determine if a set of subtables is consistent? The iteration is programmed to halt when the differences between the appropriate sums and the subtable values are all less than a margin of error, epsilon. Therefore, we could allow the computer to make many iterations, and, upon failure to halt, we could conclude that some comparable subtable values must differ by more than epsilon. Actually, we first use another algorithm to discover inconsistencies: we take every pair of subtables and compare them to find the common dimensions. Then we collapse each of the subtables to the tables corresponding to the common dimensions and compare these tables cell by cell. If we discover differences greater than epsilon, we reject one of the tables.

Now, at first blush it would seem that the procedure of making every possible cell by cell comparison would eliminate any inconsistencies in the data. However, it turns out that this is not the case. In fact, there are certain combinations of numbers and dimensionalities of subtables which can produce inconsistencies even when this cell by cell comparison has been made for every pair of the subtables. We show such a pathological example below. Here the faces of the cube represent the values of the three two-dimensional subtables. If we compare the

marginals for each of the three two-dimensional subtables, we will find that the marginals are equal; i.e., there are no obvious inconsistencies in these tables. However, it is easy to see that these tables cannot be produced by any set of eight positive interior values.

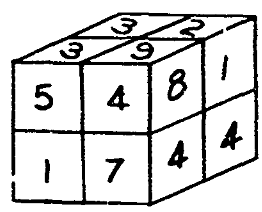


Fig. II-2. A Pathological Cube.

This is a rather amazing inconsistency. It means that social scientists can bring sets of data from various sources together which on the face, applying every reasonable test, appear to be consistent and yet have a basic inconsistency within them. It seems from our experiments with these data that this situation arises whenever the subtables are such that they completely enclose the space of the grand table. For example, in the case of the cube above, the three two-dimensional tables completely enclose the space of the cube. Any two of these tables taken together do not enclose the space of the cube and, in fact, do not produce any inconsistency. Likewise, we have found that four three-dimensional tables, which completely enclose a four dimensional space, also often generate these

inconsistencies. At one point in the programs which account for the duplication data, we must combine several three-dimensional tables, and these inconsistencies have arisen. At this point we throw out some of the tables until we arrive at a subset of the subtables which are consistent. The decision about which tables are thrown out is made on an intuitive basis by the researcher in interaction with the computer, as we have not yet been able to formulate a way to decide which table is the best table to throw out initially.

Since it is so difficult to detect these inconsistent tables, we have, in the case of the duplication routines, introduced a second attempt to find these inconsistent tables. It may be obvious to the reader that the values in the grand table are related to the values in the subtables by a series of linear equations. For example, in the case of the cube above, we can relate the values in the interior of the cube to the values in the faces by a set of seven simultaneous equations. Since these equations contain eight unknowns, there is no unique solution to the equations, but an infinite number of solutions to the equations, given that the faces of the cube are consistent. If the faces of the cube are inconsistent, however, there is no solution. Therefore, in the duplication routines we use a linear programming algorithm to find a feasible solution to the equations or to find that there is no feasible

solution to the equations. If the linear programming code declares that there is no feasible solution to the set of equations, we know that the tables themselves are inconsistent.

Thus, the requirement that the tables enclose the m-dimensional space of the grand table in order for inconsistencies to occur is somehow related to the possibility of having no feasible solution in a set of linear equations. We have not explored this further, but we would suppose that in the algebra of linear equations there is some theorem about the possibility of no feasible solution which is also interpretable as a theorem about when a set of equations encloses a hyperspace.

We offer one more observation about the relationship between the iteration technique, solutions of linear equations, and maximum entropy extensions of probability distributions. Brown's proof of convergence of the algorithm rests on the fact that at each stage of the iteration, the entropy of the resulting distribution must be larger than the entropy at the preceding stage. Therefore, one possibility for a test for consistency of the tables (or for a test of when to stop iterating) is to calculate at the end of each iteration cycle the entropy of the resulting distribution and see if it is in fact larger than at the end of the previous cycle. It would seem that at a point where the entropy has not increased, we either have reached the

optimal solution (or at any rate a solution which cannot be carried any further because of the problem of significant digits and rounding error), or we have reached the point where the inconsistencies in the tables preclude any further approximation to the final distribution.

Now consider the relationship to the set of linear equations. The iteration appears to be a new way of finding a feasible solution to a set of linear equations or of finding that there is no feasible solution. If any set of linear equations can be thrown into this form, then perhaps it would be worthwhile to compare the speed of this solution with that of the standard Simplex solution.

We have digressed for a moment to explain in detail one of the most fundamental simulation routines and a general estimation process which seems to have desirable features in several different kinds of problems. Now we turn to the actual modeling of the Cincinnati mass media system of 1947-48.

CHAPTER III

CONSTRUCTING THE MEDIA SYSTEM I: THE SIMULATION POPULATION AND THE NEWSPAPER AUDIENCES

The Role of a Population in Simulations

Some simulations--we are thinking here particularly of simulations of economic systems--relate dependent and independent variables quite directly through systems of equations. The data describing the independent variables are measured or assumed by the researcher and the implied values of the dependent variables follow rather directly from these equations. In these simulations, some body of theory rather closely relates the values of the independent and dependent variables. These simulations do not seem to be characterized by what one might call sub- or micro-elements or a population, by which we mean components of the model which are, in some theoretical scheme, at a lower level of analysis than other elements. Thus, sectors of an economy, or individual concerns, or a number of (real or symbolic) people comprising a simulation population may function at a lower level of analysis than that at which one analyzes the outputs of the simulation. These outputs may be in terms of some

net or aggregate statistics or conditions for the system, e.g., stability or instability, total income or G.N.P., or exposure; or in terms of comparisons of such aggregates among population subgroups, e.g., income and G.N.P. of sectors of the economy, or exposures of males or females; the outputs are usually not in terms of the smallest population element. This is true for the mass media simulation; the outputs which we wish to analyze are not in terms of individual members of the population but rather in terms of the various subgroups.

There are at least three reasons why it may be necessary to define elements of a simulation at a level so removed from the statistical outputs that they are never used for analysis:

1. It may be (as in the present simulation) that the input data, although aggregated over subgroups or the entire population, come in variables too diverse to be dealt with at the aggregate level, given the present state of the theory. The theory often relates to a lower level and therefore the data must be disaggregated, hung (so to speak) on some set of population points, and later recombined or integrated. Thus in the present simulations we proceed from the level of population and audience subgroups, total cumulations, and duplication,

down to hypothetical individuals with appropriately arranged exposure probabilities; we then add a driving mechanism, the message scenario, to get individual exposures, and then reaggregate to get exposures again at the level of the group or the total population. The individual points do not perform any action or interact with each other at all. We might characterize these as disaggregative-integrative simulations, or consistency machines.

2. It may happen that the macro-level theory is well developed, but that the data are available only at a lower level, therefore making it necessary to model at the lower level and then aggregate. This would also allow exploration of the effects of changes at the lower level (e.g. in a particular market) on the aggregate system.

3. The researcher may be interested in complex processes at a lower level, (partially) validating them by the resulting aggregate output. This is often the case when the population or subelements are in interaction or "coupled" with each other. Examples of this kind of simulation include the various models of small groups and

the Abelson and Bernstein fluoridation simulation.[1]

In the first two cases above, where the population is uncoupled, the researcher should consider whether he might not be able to describe the relevant population subgroups using mathematical functions, rather than with individual population points. In the present simulation this would seem to be a just feasible task; for coupled populations the mathematics are probably too complex. Nevertheless, as researchers become more familiar with modeling and more sophisticated mathematically, simulations such as this will probably be more completely described in terms of mathematical functions, with a resulting great economy of running time, increase flexibility, and ease and clarity of analysis.

In a simulation using an uncoupled symbolic population, it seems unnecessary that the distribution of the population points match either exactly or proportionately the distribution of the modeled population. In a coupled population where the data points interact,

---

[1]Abelson makes the important distinction between ccupled and uncoupled populations in his chapter on simulation for the revised Handbook of Social Psychology, Lindzey, Gardner and Aronson, Eliot (editors). In press, Addison-Wesley. A description of the fluoridation simulation is found in Robert P. Abelson and Alex Bernstein, "A Computer Simulation Model of Community Referendum Controversies," Public Opinion Quarterly, XXVII, (1963), 93-122.

there may be for some processes a "critical mass" which quite changes the behavior of the system and therefore necessitates a scale model (at least) of the real population. However, for a stochastic simulation like the present one, it may be much more efficient to use a "stratified sample"--even equal numbers in each population subgroup or cell--if we wish to do a detailed analysis of some important, but proportionately, small, subgroup. The computer is just the tool to carry out complex weightings of population subgroups in calculating aggregate statistics. Of course, if each population subgroup is described by some mathematical function rather than actually produced as individual points, then these functions carry the subgroup weightings.

With these preliminary remarks, we turn now to the actual construction of the Cincinnati mass media simulation population.

The Cincinnati Simulation Population

The simulation population consists of a certain number of hypothetical persons distributed throughout the space defined by the attributes or variables which have been selected as relevant to the process being modeled. The points in this property space (we will usually call them "cells" or "population types") represent all possible combinations of the levels or categories of all attributes. Thus, a typical population type might consist of all

young, high school educated, middle status, males who are interested in international affairs.

Since the present simulation model was not programmed to allow weighting of cells or audience types, it would seem, for the Cincinnati simulation, that we would like to replicate, in miniature, the distribution across cells of the 1947-48 Cincinnati population. However, the choice of a population base for the simulation is really not this straightforward. Actually the target populations surveyed by NORC and the Census, and by several of the other surveys to be presented below, were not identical. During the 1940 census, a certain urban area around the city formed what was known as the Cincinnati Metropolitan District. This was the basis also of the 1947-48 NORC surveys. However, for the 1950 census, which probably is a better representation of the 1947-48 Cincinnati population, the boundaries of this area were slightly changed and the new unit was called the Cincinnati Metropolitan Area. Also, the base for reporting newspaper circulation figures and for several Times-Star audience surveys was a slightly different area called the Cincinnati City Zone. Finally, the Hooper ratings used in calculating the radio audiences are produced from a sample of the telephone homes located in the non-toll-call area of the city.

In order to simplify matters, we shall ignore these small differences in the target populations for

these surveys, and turn to the obvious larger differences, the errors in the NORC sample, and (in Chapter IV) the Hooper sample for radio audiences.

The NORC sample, even after weighting, was not an accurate representation of the Cincinnati population. Tables III-2 and III-3 show the breakdown according to the September 1947 NORC sample and according to 1950 census figures for the Cincinnati Metropolitan Area. It can be seen that in several ways the sample and the population differ. People with only grade school education are significantly underrepresented in the sample and younger people are also slightly underrepresented.

Table III-1. Distributions of the NORC Sample by Sex, Age, and Education.

	Male			Female		
	Age			Age		
Education	21-39	40-	Marginal Totals	21-39	40-	Marginal Totals
College	5.72%[a]	6.35%	12.07%	3.91%	2.18%	6.09%
H.S.	11.70	7.17	18.87	16.43	10.43	26.86
G.S.	4.99	10.80	15.79	4.81	15.51	20.32
Marginal Totals	22.41	24.32	46.75 (N=515)	25.15	28.12	53.25 (N=587)

[a]The entries are percentages of the weighted sample (N=1120).

Table III-2. Distribution of the Adult Population of the Cincinnati Metropolitan Area by Sex, Age, and Education According to the 1950 Census.

	Male			Female		
	Age		Marginal Totals	Age		Marginal Totals
Education	21-39	40-		21-39	40-	
College	4.17%[a]	3.03%	7.21%	3.16%	2.55%	5.71%
H.S.	9.73	6.89	16.63	13.25	8.80	22.05
G.S.	6.33	16.85	23.18	6.50	18.70	25.20
Marginal Totals	20.23	26.77	47.02	22.91	30.05	52.96

[a]The entries are percentages of the total population (N=617,548).

Table III-3. Percentage of the March, 1948 NORC Sample Reporting Exposure to at Least Three Media, by Education, Age and Sex.

Dimension	Percentage Exposed to At Least Three Media
Education	
College	68
High School	43
Grammar School	17
Age	
21-39	44
40-	32
Sex	
Male	43
Female	34

These differences are troublesome for the following reasons. On the one hand, the surveys, circulating figures, etc., by which the audiences and eventual exposures are calculated, are presumably valid for the Cincinnati population as described by the 1950 census. On the other hand, after processing all the data through the simulation, we would like to correlate the theme exposures of the simulated population with the changes in information and opinions which were measured by the NORC survey. Thus, if we model the population as recorded by the census, the circulation and audience statistics are presumably directly valid as input, but the output must be interpreted for purposes of comparison with the panel data as rates of exposure (or percentage exposed) for each subgroup. Conversely, if the survey panel is modeled in the simulation, then the input data must be modified so as to be valid for this different population configuration. This is the course which has been followed here. Since much of the data used to establish newspaper audience ratings for the population types is provided by the NORC survey itself, and since audience data for the radio is known only by sex--a breakdown correctly reproduced by the NORC weighted panel--we have used the weighted NORC data for construction of the simulation population. Essentially, the audience input data is so poor that it hardly justifies constructing a simulation population exactly modeling that of the Cincinnati Metropolitan Area.

The dimensions chosen to describe the simulation population were sex, age, education, socioeconomic status, and initial (September) level of interest in international affairs. All of these dimensions were recorded in the NORC survey. In an entirely urban population, these dimensions seemed the most critical for an explanation of exposure to international news. For example, the report on the Cincinnati Plan found that

> Among those rated as 'interested' in September by virtue of their having expressed 'keen' interest in two or more of four given international topics, 47 percent reported exposure to three or more of the media by March; whereas, among those relatively 'uninterested,' 29 percent did.[2]

The March survey also showed the effect of education, age, and sex on media exposure (Table III-3).

Table III-4 indicates the levels and marginals for each of these dimensions. This combination of dimensions and levels divides the population into 144 population types.

The NORC September survey before weighting was a sample of size 745. In weighting to match the population, many respondents' cards were duplicated so that the resulting weighted sample size was 1120. Dividing this number by the 144 audience types gives an average of only 7.8 weighted panel members per audience type. With cells both well above

---

[2]The quotation and table are taken from Shirley A. Star and Helen MacGill Hughes, Report on the Cincinnati Plan, p. 9.

Table III-4. Marginal Percentage Breakdowns from the NORC Survey for the Dimensions of the Simulation Population.

Dimension	Percentage Marginal
Sex	
Male	46.8
Female	53.2
Age	
21-39	47.7
40-	52.4
Education	
College	18.1
High School	45.7
Grade School	36.1
S.E.S.	
High	19.2
Middle	62.2
Low	18.7
September International Affairs Interest	
Highest	25.9
Upper Middle	22.0
Lower Middle	28.1
Lowest	24.0

and below this average, we obviously are estimating the size of many cells from a very small sample. Thus, this is just the situation to use the Mosteller parameter estimation technique. This we have done, constructing the simulation population from the 10 possible <u>three</u>-dimensional subtables.

Finally, in order to reduce the variance due to the stochastic procedures of the simulation (e.g., the random assignment of probabilities within cells) we have chosen the size of the simulation population to be 2,000 computer individuals distributed across the 144 audience types.

The Vehicle Audience Distributions

Having constructed the simulation population, we now estimate the breakdown across the population types of the average audience of each of the vehicles. For each of these types, the number in the vehicle audience divided by the total number of that type provides a mean probability of exposure or audience <u>rating</u> for the vehicle and audience type. First we turn to the newspapers.

The Cincinnati Newspaper Audiences

There were in Cincinnati at the time of the NORC Survey three daily newspapers, the *Enquirer*, the *Times-Star*, and the *Post*. The *Enquirer* was a morning newspaper; the *Times-Star* and the *Post* were evening newspapers. In addition to the Monday through Saturday issues of each of these newspapers, the *Enquirer* published a Sunday newspaper which

we have included as a fourth newspaper vehicle. Thus we wish to estimate the distribution of the average audience of each of these four vehicles over the 144 possible audience types.

The NORC survey asked about each respondent's daily reading habits and found the fundamental data from which each audience distribution has been estimated. The question asked of each respondent was "Do you usually read a daily newspaper?"; if "yes," "Which?". The readership found for each of the daily newspapers was: for the Enquirer, 43 percent; for the Post, 52 percent; and for the Times-Star, 55 percent. Four percent of the population reported reading any other newspaper, and three percent read no daily newspaper. Because of the wording of the question these audience percentages are almost surely slightly high. Generally, the most accurate method of measuring the average occurrence of a frequently recurring event is to identify a fixed period of time for the measurement.[3] In media research the tactic is to ask about the exposure to the last issue of the vehicle; for daily newspapers this means asking about exposure to yesterday's (or Saturday's) issue. Thus

---

[3]If there is a significantly large proportion of the population which is exposed to about half of the issues and whose answer to the question as phrased below is likely "yes," then the resulting percentage will significantly overestimate the average audience.

the standard newspaper readership question is of the form "Which daily newspaper or newspapers did you read yesterday?"[4]

Not only are the NORC readership figures for each newspaper probably too high, but the form of the question probably also causes an underestimation of multiple newspaper readership. However the disproportionate sampling of the higher educated, higher status members of the population, i.e., of those who are more likely to read more than one newspaper, may compensate to some degree for the bias introduced by the wording of the question.

Because of the problems of the NORC data, the small sample size, the method of drawing the sample, and the phrasing of the question, it would be helpful to have additional evidence about the newspaper exposure. This evidence is available from two audience surveys conducted by the Cincinnati Times-Star, one in 1947 and the second in 1951. These were essentially quota samples of housewives living in the Cincinnati City Zone.[5] The question asked was

---

[4]In the aided recall method, copies of each of the previous day's local newspapers are shown to the respondent. This aid usually produces a somewhat higher readership.

[5]These surveys are described in detail in The Flow of Retail Buying Traffic in Cincinnati, 1957: Fifth Survey, The Cincinnati Post, (Cincinnati, Ohio: 1947) and Buying Habits Survey: The Cincinnati Market (1951), the Cincinnati Post (Cincinnati, Ohio: March 1952). The samples were quota samples of all housewives in the Cincinnati City Zone, controlling for economic groups, geographic location, race, and types of residence. The sample size for the 1947 survey was 2500 and for the 1951 survey was 4588.

"Which daily newspaper or newspapers were read in your home yesterday?" This phrasing of the question probably also biases the readership upward insofar as every adult member of the family was not a reader of the paper. The data for the three surveys are presented in Tables III-5 and III-6 below.

Table III-5. Total Readership of Each Daily Newspaper According to Three Surveys

	Total Readership		
Survey	Enquirer	Times-Star	Post
Times-Star, 1947	37.44%[a]	54.84%	51.12%
NORC, September, 1947	43.50	55.30	52.40
Times-Star, 1951	42.70	48.85	48.43

[a]Data represent percentages of the adult population.

In examining these tables, we probably should discount somewhat the data from the Times-Star survey of 1951 since we might reasonably expect some change in readership habits over the four-year interval from 1947 to 1951. For the two 1947 surveys we find reasonably close agreement except for two cases, first for the readership of all three newspapers for which the NORC survey finds approximately twice the readership (14.12%) of the Times-Star survey and second, for the average audience of the Enquirer, which the

Table III-6. Combination Readership of the Three Daily Newspapers According to Three Surveys

Survey	Readership of the Daily Newspaper Combinations			
	Enquirer Only	Times-Star Only	Post Only	Enquirer and Times-Star
Times-Star, 1947	6.12%[a]	25.[illegible]2%	27.12%	14.72%
NORC, September 1947	7.88	24.30	23.85	11.91
Times-Star, 1951	10.48	22.54	27.29	15.48

Survey	Readership of the Daily Newspaper Combinations				
	Enquirer and Post	Times-Star and Post	Times-Star, Post and Enquirer	None	Total
Times-Star, 1947	9.60%	7.40%	7.00%	2.52%	100.00% (N=2500)
NORC, September 1947	9.62	5.04	14.12	3.30	100.00% (N=745)
Times-Star, 1951	10.31	4.40	6.43	3.07	100.00% (N=4588)

[a]Data represent percentages of the adult population.

Times-Star survey finds to be about 37.4 per cent, and which the NORC survey finds to be 43.5 per cent. The first discrepancy, the finding that the triple readership is larger in the NORC survey than in the Times-Star survey, comes somewhat as a surprise, since the question phrasing for the NORC survey would seem to bias the triple readership downward. Perhaps the discrepancy occurs because of the biases in the NORC sample although it is difficult to imagine that the NORC probability sample, even though it is a smaller sample than the Times-Star sample, was less accurate a representation of the population than the Times-Star quota sample.

To summarize our feelings about the data at this point, we would say that the average audiences are probably somewhat too high because of the phrasing of the survey questions and this holds for all three surveys. For the NORC survey it seems reasonable to suspect that the measurement of the triple readership is significantly too high.

We do have the results of one other survey which somewhat corroborates the two findings in the paragraph above. This study, A National Study of Newspaper Reading: the Functions of Newspapers for Their Readers, was conducted during March and April, 1961, by the Audits and Surveys Company from a national sample of newspaper

readers.[6] In this study the combination of two samples upon which most of the data is based totaled 2,449 households, with 4,826 individual respondents. From a national survey we are not able to ascertain the readership of individual newspapers, such as the Cincinnati newspapers; however, we do have data on the number of newspapers read by various percentages of the population. (Of course, when we look at this data, we must remember that it is national data taken during the year 1961, that is, fourteen years after the time of the other surveys.) Table III-7 below shows readership by number of newspapers for each of the four surveys.

Table III-7. Average Multiple Newspaper Readership in Four Surveys

Number of Newspapers Read	Survey			
	Times-Star 1947	NORC 1947	Times-Star 1951	National Study 1961
0	2.52%[a]	3.30%	3.07%	20.3%
1	58.76	56.03	60.31	53.1
2	31.72	26.57	30.19	22.5
3 or more	7.00	14.10	6.43	4.1
Total	100.00% (N=2500)	100.00% (N=745)	100.00% (N=4588)	100.00% (N=4368)

[a]Data are percentages of the population.

[6]Audits and Surveys Company, Inc., A National Study of Newspaper Reading: the Functions of Newspapers for Their Readers (New York: Audits and Surveys Company, Inc., 1961).

The 1961 national study found generally smaller multiple readership and generally higher non-readership of the average daily newspaper. Probably most of the difference between the 1961 national study and the other studies is due to changes in readership during the intervening fourteen years. By 1961 there were far fewer three-newspaper cities and, in addition, it is well known that newspaper readership has not kept pace with the growth of the population in the last twenty years. However, the 1961 study does add another piece of evidence which seems to indicate the possibility of an overestimation of readership on the part of the first three surveys.

There exists one more piece of evidence which seems to indicate that the estimates of readership according to the 1947 NORC survey are somewhat high. From the 1950 census we find the number of adults living in a household in Cincinnati was, on the average, about 2.15 adults per household. (There are two ways to calculate the number of adults per household, depending essentially on the definition of a household. The two calculations give a minimum number of adults per household as 2.12 adults per household and a maximum number of 2.23 adults per household.) The 1961 readership study found 2.04 adult readers per newspaper copy. This figure seems quite suggestive if there were, nationally, about 2.15 adults per household. One might conceive of almost every copy of a newspaper sold

going into a household and, in general, the average adult in the household reading the newspaper. From the city zone average circulation of Cincinnati newspapers for the year ending March 31, 1948, and the NORC audience estimates, we have calculated the number of adult readers per copy of the newspaper. For the three newspapers, these figures come out remarkably close together. For the Post and the daily Enquirer, the estimate gives 2 [illegible] adult readers per copy; for the Times-Star, the figure is 2.58 adult readers per copy of the newspaper. Although these figures are very close, they exceed substantially the figure of 2.04 adult readers per copy found in the 1961 survey. Moreover, they also exceed the estimate of 2.15 adults per household in Cincinnati in 1950. Thus this bit of evidence also indicates the possibility of the NORC survey's overestimation of the average newspaper readership. (Of course, we must realize that the adult readers projected from the NORC survey are not the ordinary adults from the Cincinnati population since the NORC survey is not a faithful reflection of the Cincinnati population. The NORC adults are more likely to read newspapers, and therefore we might expect the number of adult readers per copy to be higher among this group than among the population of Cincinnati as a whole.)

To sum up, it seems hard to believe that on the average day only 3.3 per cent of the NORC population fails

to read at least one newspaper. Therefore, we have somewhat arbitrarily decided that this figure should be increased to ten per cent of the adult population.[7]

The correction is not as simple as it would first appear. The procedure finally decided upon is as follows: we assume that everyone who answered "no" to the question by readership of a newspaper did indeed not read the newspaper. However, we shall assume that a constant proportion of those people who did answer "yes" to the question of newspaper readership were not in fact in the average audience of the newspaper. This constant proportion then represents a probability of not being in the newspaper's average audience given that the respondent answered "yes" to the question about newspaper readership. Then, for example, the number of people in the group who claimed to read three newspapers, but who in fact on an average day read no newspapers, is the number claiming multiplied by the cube of the probability. Likewise, the number who claimed to read two newspapers but in fact read no newspapers on an average day is the claiming number multiplied by the square of the probability. If we let K_i represent the percentage claiming to have read i newspapers (i = 0, 1, 2, or 3), we arrive at the following cubic equation which must be solved for q, the probability of not reading, given a "yes" answer:

[7]The 1961 national survey also found in cities of 500,000 or more than 16.1 per cent of the adults were nonreaders of a newspaper on an average day.

$$10\% = K_3q^3 + K_2q^2 + K_1q + K_0$$

or

$$10 = 14.13q^3 + 26.60q^2 + 56.06q + 3.21 \ .$$

An approximate solution to this equation is a value for q equal to 0.1145. Therefore, the probability that a person did read a newspaper on an average day, given that he claimed readership, is 0.8855. This is the figure which gives an average audience, exposed to at least one newspaper, of 90 per cent of the population. Therefore, in constructing the audience tables for the simulation, we have used the audience tables given by the NORC survey, however, multiplying every figure given by the NORC survey for readership by the factor 0.8855.

Just as in the case of the population, the tables actually entered into the computer from the NORC survey were the ten three-dimensional tables which can be constructed from the five dimensions of the population. The rationale for this was exactly that of the population; namely, that the sampling errors from such a small sample would not justify entering in the full five-dimensional table. Therefore, the iteration technique was used on the audience as well as on the population. After the tables were entered into the computer, the correction factor spoken about in the paragraph above was entered into the computer to multiply each of these tables before the iteration was

performed. The final average audiences for each of the three daily newspapers resulting from the correction are as follows:

Average Audiences as Percentages
of the Total Population

Daily *Enquirer*	*Times-Star*	*Post*
39.2%	50.2%	47.6%

The NORC survey asked only about the readership of the daily newspapers; there was no readership data for the Sunday *Enquirer*. Likewise, the *Times-Star* surveys did not ask about readership of the Sunday *Enquirer*. Therefore, we have had to make educated guesses about the distribution and the size of the audience of the Sunday *Enquirer*. As a base line for estimating the average audience, we take the average circulation in the city zone through the year ending March 31, 1948. This circulation was 180,150 or about one and one-half times that of any of the daily newspapers. The 1961 national survey found that the number of readers per copy of the Sunday newspapers was slightly lower than that of the weekday papers. For the Sunday newspapers, the study found 1.92 adult readers per copy, compared with 2.04 for the weekday newspapers. Therefore, we have three possible estimates of the readership of the *Enquirer*; the lowest estimate is derived from the value of 1.92 adult readers per copy, the highest estimate from the 2.59 adult readers per copy, estimated from the NORC

data for the daily newspapers. We have chosen a middle range estimate, that estimate which is produced by using the number of adults per household as the number of adult readers per copy. This figure, 2.23 adults per copy, produced an average readership for the Sunday Enquirer of about 65 percent of the population.

Having estimated the average audience, however, we still must estimate the distribution of the audience throughout the population. What we have done is to assume that the audience is distributed as is the audience of the daily Enquirer, making, however, some interesting changes in the iteration technique once we have made this assumption. The ten three-dimensional tables that describe the audience of the daily Enquirer were entered as tables for the audience of the Sunday Enquirer. However, the figures in these tables were each multiplied by the ratio of the average audience of the Sunday Enquirer to the average audience of the daily Enquirer in order to increase the figures to produce the 65 per cent average audience for the Sunday Enquirer. Of course, if some cells of the population are nearly at the saturation level for readership of the daily Enquirer, this increase may very well make the audience of these cells larger than the population of the cell. We must in some way account for the distribution of the population as we iterate upon the audience figures for the daily Enquirer; therefore we have

initialized the values of the iteration technique with the population values as derived in the previous iteration. Thus the daily Enquirer inflated audience values are biased by the population cell values to produce the cell values for the Sunday Enquirer audience. This method is not foolproof, and it still is possible for some cells to have audiences greater than the population for these cells. Of the 144 cells in the population, eighteen cells were somewhat larger in audience than in population. These cells were then shown to the researcher by the simulation and the researcher was asked to pick appropriate probabilities of exposure, i.e., average audiences for the inconsistent cells. After correcting the 18 improper cells by setting their cell means at a value of .95, the simulated audience for the Sunday Enquirer was 64.3 per cent of the population, remarkably close to the 65 per cent estimated from the circulation.

Thus, we have produced the cell-by-cell mean audiences for the four Cincinnati newspapers in the Cincinnati mass media system. The resulting average audiences as proportions of the population for the four Cincinnati newspapers are shown below.

Final Average Audiences for the Four Cincinnati Newspapers As Percentages of the Total Population

Daily Enquirer	Times-Star	Post	Sunday Enquirer
39.2%	50.2%	47.6%	64.3%

CHAPTER IV

CONSTRUCTING THE MEDIA SYSTEM II: THE RADIO AUDIENCES

The second important source of United Nations and international news during the six-month period was radio. In the NORC panel at the end of the period, 53 percent reported hearing radio news programs about the United Nations and 26 percent had heard short radio mentions of the United Nations between programs.[1] Unfortunately, no relevant record of radio news messages during the period has been found (probably nothing exists comparable to the newspapers themselves which, of course, are available in library microfilms). Therefore, we have had to rely on educated guesswork plus some press comment and the memories of those who participated in the campaign twenty years ago.

At the time of the study, five AM stations, two FM stations, and one television station were broadcasting in Cincinnati. For our purposes, the effect of the two FM stations and the television station were negligible. According to the City Hooperatings, the largest combined

[1]The complete breakdown by information source is found in NORC, Cincinnati Looks Again, p. 27.

rating of all radio stations other than the five AM stations in Cincinnati was 1.2, i.e. 1.2 percent of Cincinnati radio homes at most were ever listening to radio stations other than the five AM stations.[2] The mean rating was approximately 0.5.

Likewise, the television audience was very small at this time, not because of lack of interest, but simply because there were so few television sets in service. According to Variety, in the fall of 1948 (nine months to a year after the period of the study) there were only about 4000 television homes in Cincinnati.[3]

Therefore, we shall restrict our attention to the audiences of the five AM stations. We shall assume that by far the largest volume of United Nations and international news broadcast during the period was through the regularly scheduled news broadcasts. The times of these broadcasts have been determined from the daily radio schedule carried in the newspaper. (The content relevant to the themes to be studied will be inferred from the front page news of the newspapers for the day. This is described in more detail in Chapter VII.)

---

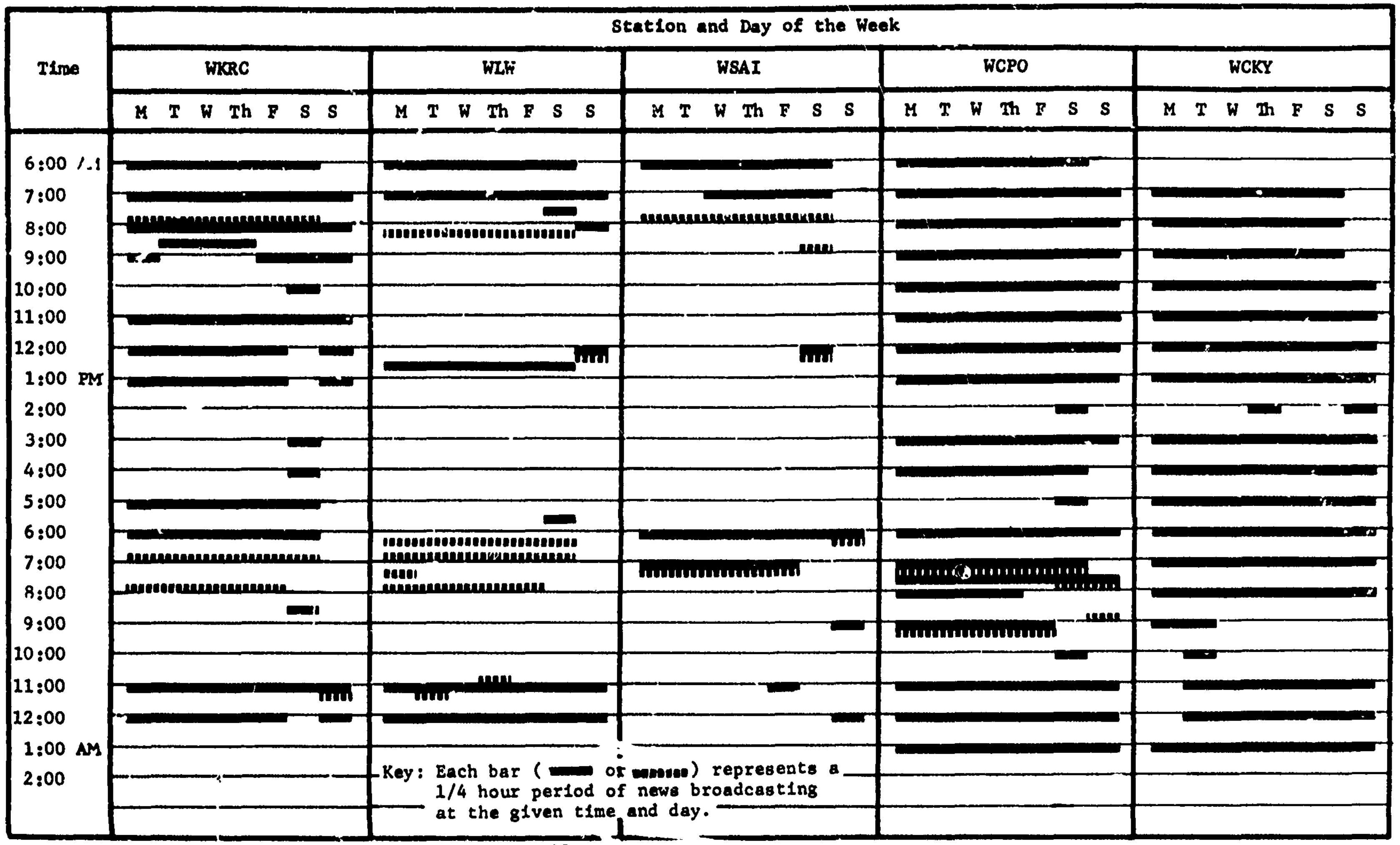
[2]This rating is an average through the months October, 1947-February, 1948. This peak rating occurred Sunday afternoons at 12:30 and 4:30 P.M. See C. E. Hooper, Inc. City Hooperatings: Cincinnati, Ohio, Fall-Winter Report: October, 1947-through February, 1948 (New York: C. E. Hooper, Inc., 1948).

[3]"U.S. Video Sets Now 484,350," Variety, August 25, 1948, p. 71 found one television station and 6,000 television sets in circulation in the 40-mile service area of Cincinnati.

These broadcasts were usually fifteen minutes in length, but also include the five minute newscasts inserted at the beginning of the hour in a continuing disk jockey program, and in addition, several "name" newscasters or commentators who usually broadcasted for fifteen minutes in the evening including H. V. Kaltenborn, Lowell Thomas, Edward R. Murrow, Gabriel Heatter, Elmer Davis, etc. The distribution of these newscasts by station, day of the week, and hour is shown in the chart in Fig. IV-1. The chart is constructed from newspaper radio schedules for the week of the second through the eighth of January, 1948.[4] Obviously, such a variety of news programs will vary considerably in content; therefore, the selection of themes broadcast (inferred from the newspaper content analysis) must represent some sort of "average" content.

---

[4]Some ambiguities exist in this chart, of course, since there are various kinds of news broadcasts. Several of the men mentioned above, e.g. H. V. Kaltenborn and Elmer Davis, commented upon, rather than reported the news. Their commentary would not be expected to follow the daily (front page) news as closely as the reporting in an ordinary newscast. There was also a problem in identifying several of the names; Joseph Garretson, I believe, was a columnist for the Cincinnati Enquirer and could be expected to report occasionally on efforts of local United Nations groups; however, Waite Hoyt, who was for many years the broadcaster for the Cincinnati Reds, would not likely report any relevant news. Similarly, Dallas DeWeese reporting on market news at 7:30 A.M. over WLW would not be likely to report any international or United Nations news. Finally, the newspaper schedule is constructed with 15 minute time intervals; therefore, every program noted appears as at least of a quarter-hour length. For convenience, they are all so represented on the chart. However, in the Hooper rating (Hooper, City Hooperatings) there is an asterisk throughout indicating that all but four of the WCKY newscasts indicated on the charts, and several of the WCPO newscasts were actually five-minute inserts at the beginning of the hour.



Source: <u>Times-Star</u> radio schedules, January 2-8, 1948.

Figure IV-1. News Broadcasting by Cincinnati AM Stations, Week of January 2-8, 1948

There exists some independent validation of the data in the chart as well as information about the time devoted to international news. According to a Times-Star article of November 21, 1947, the Xavier University International Relations Club, in a week-long survey of the Cincinnati stations, found that

> . . . local stations were on the air for 124 hours or 7440 minutes (average per station) during the week in question. There were 81 newscasts per day for 1,005 minutes. International relations news occupied 224 minutes daily, and United Nations news rated 56 minutes, on the seven stations studied.[5]

The chart in Fig. IV-1 (of 5 stations, not including the 2 FM stations) records 442 news broadcasts for the week or 63 per day. In addition, the two FM stations combined averaged 16 news broadcasts daily for a total of 79 per day for the week of January 2-8. This compares with the 81 newscasts per day found by the Xavier study. If each news broadcast lasted 15 minutes except for the eleven on WCKY and WCPO daily, which lasted 5 minutes each,[6] then the total number of news broadcasting minutes per day was approximately 1075 minutes (cf. 1005 minutes in the Xavier survey). We note also that Xavier survey implies an average of 2.8 minutes per broadcast devoted to international

---

[5]"Students Study Local United Nations Programs," Times-Star, November 21, 1947, p. 45.

[6]See footnote 1 above.

relations news and just seven-tenths of a minute average to United Nations news.

The Definition of Vehicles for Radio

How are we to aggregate these many broadcasts into vehicles of the several themes? At first the obvious choice, seemingly analogous to the definition of each newspaper as a vehicle (despite the several editions), would be to define the station itself as the vehicle. However, we should examine more closely the meaning of the term vehicle in the context of the simulation.

What exactly is a vehicle? A vehicle is a framework for a set of messages. The framework (although not necessarily the particular set of messages) appears regularly and what is more important, it tends to have similar (not completely random) audiences, both in size and composition, at each appearance. Thus, an average audience and its breakdown along several dimensions can be defined (at least in principle). In the simulation, each member of the population has a fixed probability of exposure to the vehicle. The fluctuations in the audience size and composition (and thus in the cumulation and repeat exposure over several time periods) follow from these probabilities, i.e., since most of the probabilities are not identically 0.0 or 1.0, the variance of the audience is not 0.0. These random fluctuations will, however, in no way account for the fact that a mid-morning news message

of a given station reaches mostly women, while an evening news message of the same station reaches almost equal numbers of men and women. But haven't we provided the possibility of modifying the vehicle exposure probabilities with message exposure probabilities and format weights at a later stage in the simulation? This is true, but the effect is not the same as changing the audience composition for the vehicle. The format weight applies equally to each audience type for the particular message. The message exposure probabilities are intended to vary for different population types; however, they are constant for every message of the theme.[7] Now we shall see that during the day the size of the audience of a station varies greatly, reaching peaks at 8:00 and 9:00 A.M., at noon and at 1:00 P.M. and especially during the evening. Moreover, the composition of the audience changes as well; during the day, women are a much greater proportion of the audience than during the hours before 9:00 A.M. or after 6:00 P.M. It is therefore impossible for the simulation to match the real audience, if we treat the station as the vehicle and the several newscasts during the day as separate messages carried in the vehicle.

There is one other difficulty in the station-as-vehicle formulation, namely, the duplication. In the

---

[7]See the discussion of message exposure probabilities in Chapter VII for a more technical explanation of the format weight and message probability effects.

simulation, by far the largest part of duplication between two vehicle audiences is accounted for by the cell means, e.g. if each of two vehicles appeals more to the highly educated than to the less educated, the duplication will be larger than the random intersection of the two average audiences, since both vehicles will tend to have high probabilities of exposure in the same subset of cells or audience types. In fact, the simulation can only make small modifications from this "audience-type-produced" duplication when this duplication departs from empirical data.[8] However, in the event of two or several newscasts at the same time over different stations, because of the nature of the medium, i.e. the message exists only for a brief time and then is gone, the audience duplication will be virtually zero, even though the audience characteristics of each station be similar. Were we to use stations as vehicles then, we would find ourselves in the unhappy position of vainly attempting to simulate nearly random duplication between one pair of messages over Station A and Station B when they occur at different times, and then zero duplication between another pair of messages over the same two stations when they occur at nearly the same time.

The solution to these problems is to define the time period as the vehicle. It appears a reasonable

---

[8] See the discussion of duplication in Chapter VII.

assumption that for a given time slot and day of the week, the audience total over all stations is relatively constant both in size and composition. The steps in calculating the audience distributions for these vehicles are more or less dictated by the (poor) form of the presentation of data current at the time. The most widely distributed data are program ratings, which are actually average percentages of the non-toll-call telephone homes of a city which have a radio receiving the given program. We shall make the assumption that these ratings can be projected to all households in the population (and, of course, that the entire population lives in households). We further assume that in every household only one radio is playing at any one time, and then use listener densities, i.e. numbers of adult males and females listening per radio set playing (again, projected from the sample of non-toll . . . etc.), to calculate actual radio audiences. Finally, we solve two simultaneous equations involving these audiences and additional data to produce audiences which take account of the substantial listening outside of the home, e.g. while visiting or while commuting in cars.

"Hooperatings"

The first step is to calculate the rating for news for each hour of each day. The Hooper ratings[9] are available

---

[9]For further discussion of the method of measurement and significance of these ratings, see the Appendix on Radio Audience Measurement.

as weekly averages for each 15 minute time period from 8:00 A.M. to 7:45 P.M. These are averages for the 15 minutes over the five weekdays (Monday through Friday) and also over the five months October, 1947, through February, 1958. Although the radio audience increases from fall to winter, we shall use these average ratings over the period. Hooper provides daily half-hour ratings for the evening hours (from 8:00 P.M. to 10:30 P.M.) but these have also been averaged (for news broadcasts) to produce an average weekday since the ratings for the news broadcasts do not fluctuate greatly during the week.

As the chart in Fig. IV-1 clearly shows, the news broadcasts for a given station occurred regularly throughout the weekdays, i.e. if WKRC broadcasted news at 7:00 A.M. Monday, there was usually a 7:00 A.M. news broadcast on each of the other four weekdays. Thus the rating for one hour of a weekday suffices for the other weekdays, since the scheduling was quite constant. Indeed, as was pointed out above, the Hooper ratings are averages for the weekdays during the daytime and early evening hours. Occasionally there occurred slight changes in scheduling from day to day: sometimes these were ignored in calculating the news ratings (e.g. the 7:15 P.M. news over WLW which occurred only on Monday), or when possible the changes were averaged into the schedule (e.g. the 8:30 A.M. news over WKRC on Tuesday, Wednesday, and Thursday was considered equivalent to the

9:00 A.M. news on Monday and Friday and a constant 9:00 A.M. WKRC news was added in the ratings).

The news ratings for each hour for the weekdays from 8:00 A.M. to 12:00 P.M. (midnight) are the sum (across stations) of ratings for each news broadcast at (or quite near) the given hour. For example, the rating for the weekday 1:00 P.M. news broadcast is the sum of ratings for the following programs: 15 minutes of WKRC and WCPO news at 1:00 P.M., 15 minutes of WLW news at 12:30 P.M., and 5 minutes of WCKY news at 1:00 P.M. (WSAI broadcast no news at or about this hour). The rationale for adding these ratings is that we have no way of distinguishing the content of one of these news programs from another; therefore, we assume that everyone who listened at this hour was exposed to the same news message and that the duplication among the audiences of the programs is nearly zero. It seems unlikely that someone listening to the 1:00 P.M. WKRC news would also have listened to any of the other news broadcasts at that time.

By this process we have calculated the ratings for news broadcasts from 8:00 A.M. through 10:00 P.M. Monday through Friday, from 6:00 P.M. through 10:00 P.M. Saturday and from 12:00 A.M. (noon) through 10:00 P.M. Sunday. However, for the Saturday daytime ratings only summary data are provided by Hooper. We have estimated these ratings in the following manner: the Hooper data show that the

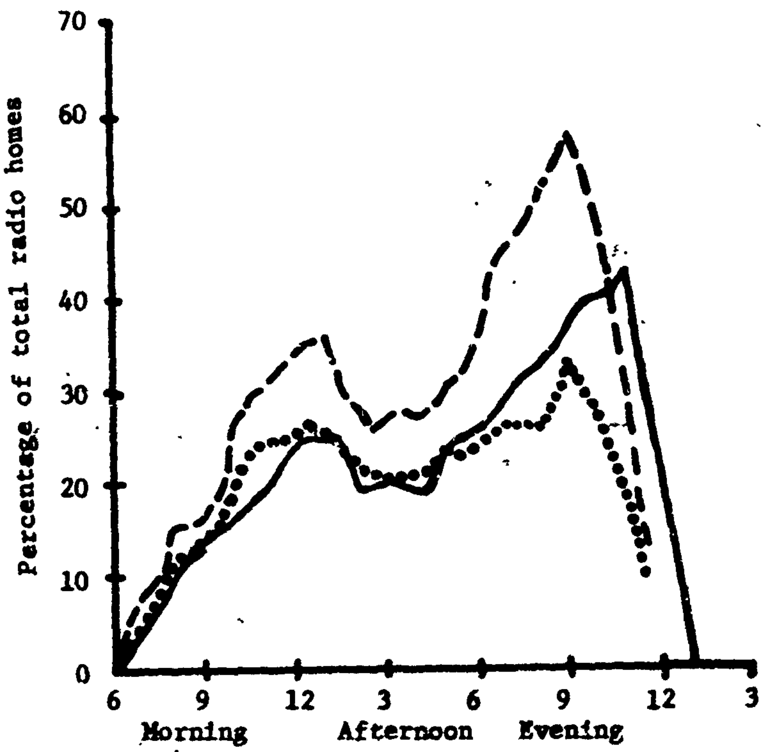
average percentage of "sets-in-use" during the Saturday daytime was 20.2 per cent, nearly equal to the average of 18.3 per cent "sets-in-use" for weekday afternoons.[10] Therefore, we have simply used the hourly "sets-in-use" averages for a weekday (see Fig. IV-2) with the "share of audience" averages for Saturday and, of course, the Saturday news broadcasting schedule, to compute news broadcast ratings for Saturday. For the Sunday morning hours 8:00 A.M. through 11:00 A.M. we have made the same calculation and used two-thirds of the resulting rating to account for families sleeping later in the morning and/or attending church, etc.[11]

The calculations above leave ratings yet to be estimated for the morning hours 6:00 and 7:00 A.M. and the evening hours 11:00 P.M. and 12:00 P.M. and 1:00 A.M. The graphs below in Figs. IV-2-7 give some indication of amount of listening during these hours. The first graph displays

---

[10]These data are taken from Hooper, City Hooperatings, p. 7.

[11]Since there was a total of only 10 news broadcasts during the hours 8:00 A.M. through 11:00 A.M. Sunday morning (compared with 15 news broadcasts during the same hours on Saturday), these ratings are not so critical.



— — — Average for four-week period, Dec. 27, 1942 - Jan. 23, 1943 (Nielsen)

.... Average for four-week period, July 11 - August 7, 1943 (Nielsen)

———— Cincinnati Weekday Average, October 1947 - February 1948 (Hooper)

Source: see note 13 above.

Figure IV-2. The Relative Number of Sets in Use (Nielsen); or, Telephone Households Listening (Hooper) (from 6:00 A.M. to 1:00 A.M.).

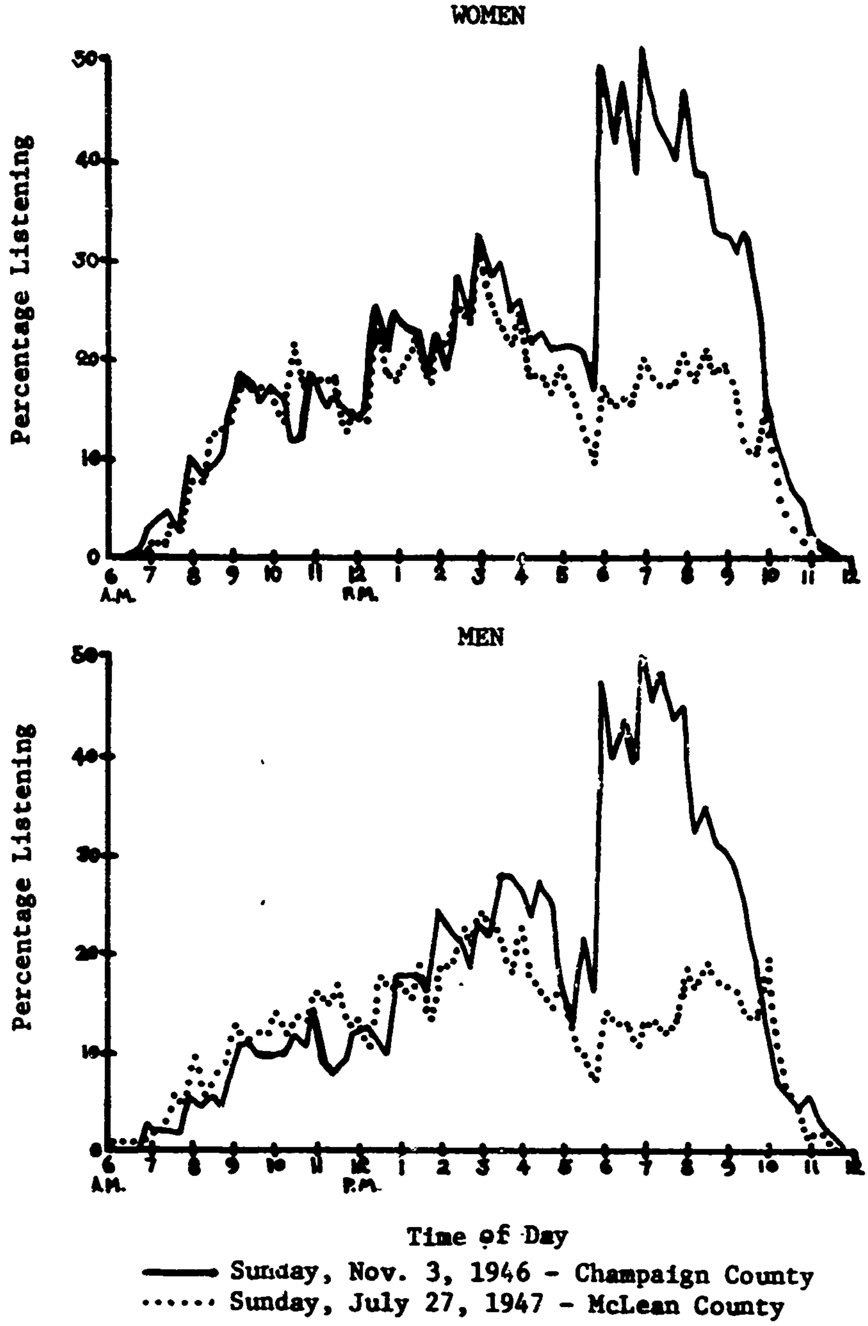


Figure IV-3. Pattern of Sunday Listening by Urban Men and Women.

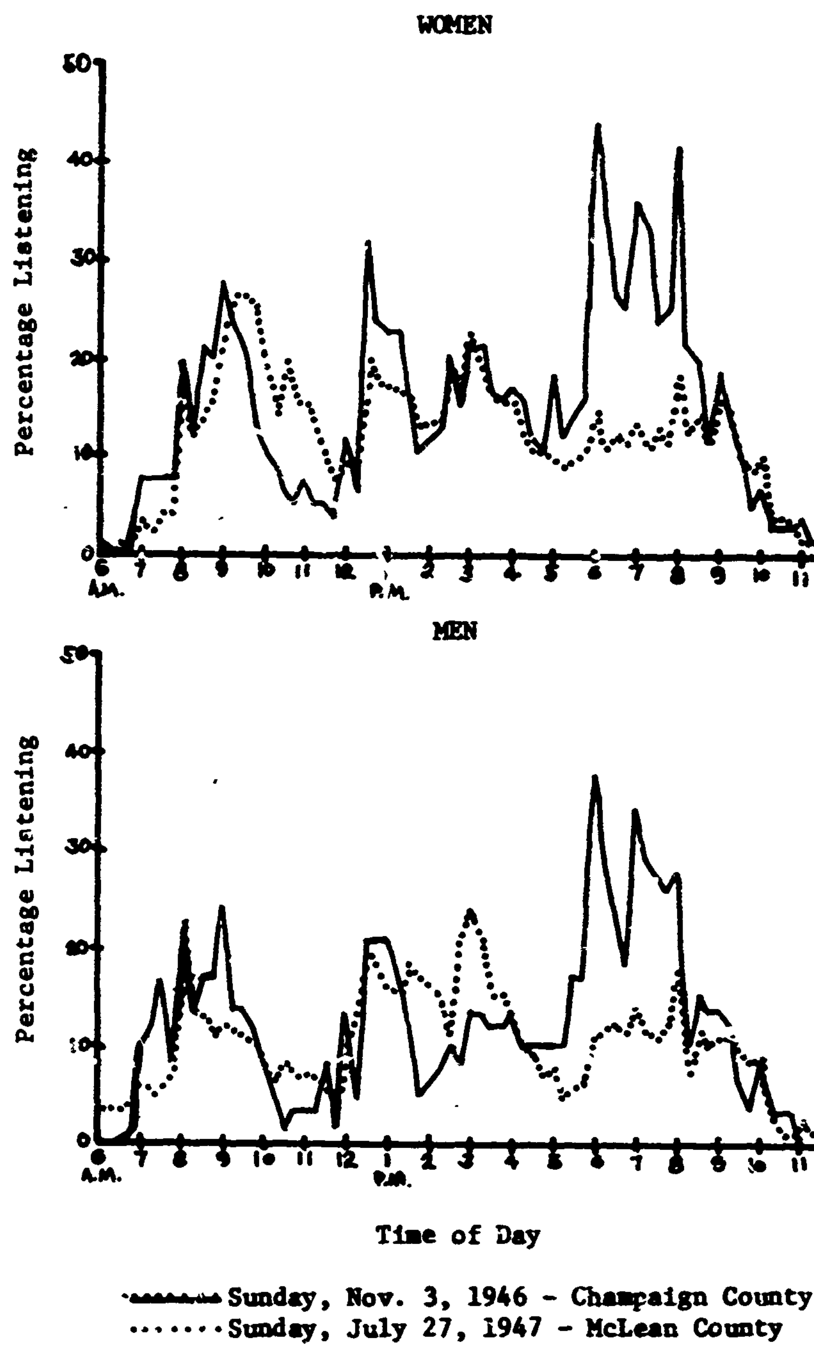


Figure IV-4. Pattern of Sunday Listening by Farm Men and Women.

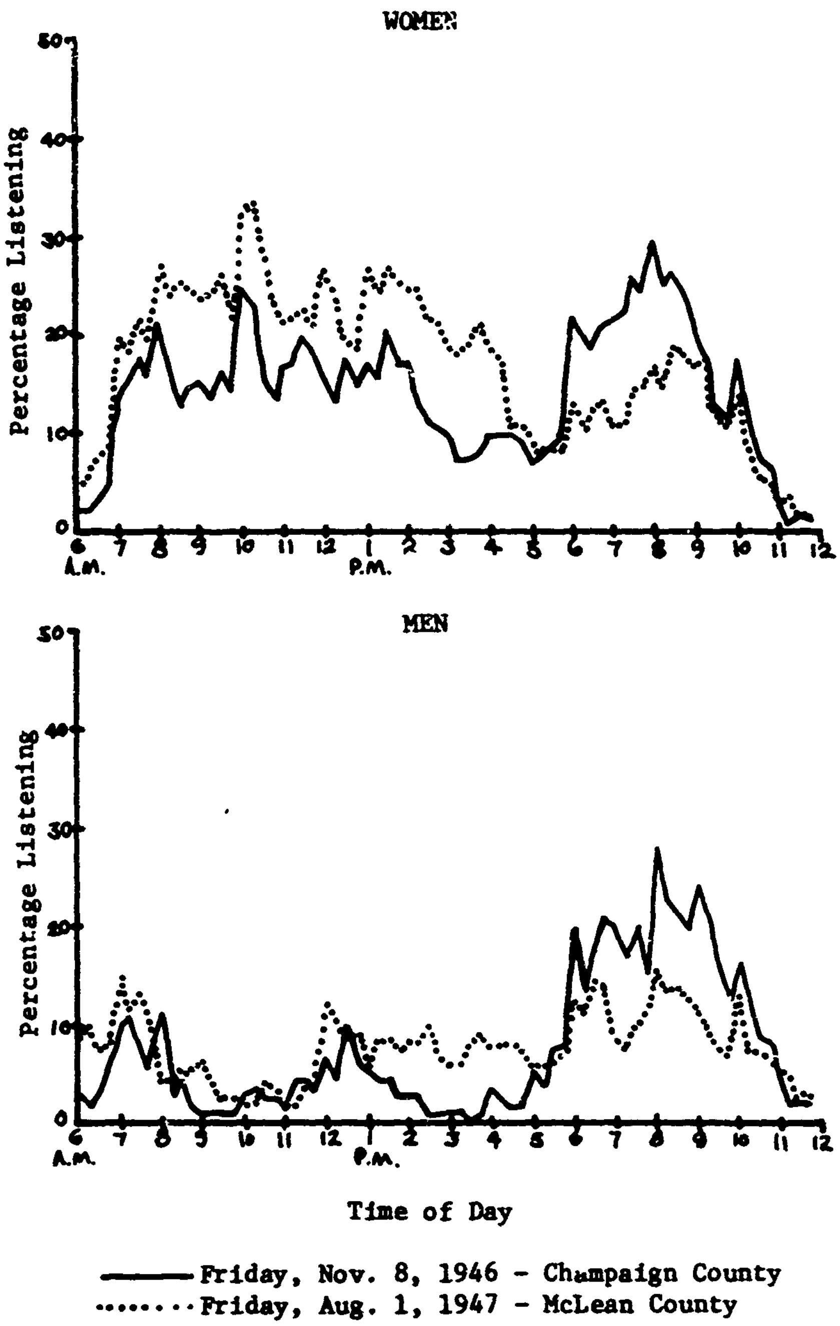


Figure IV-5. Pattern of Listening for a Weekday by Urban Men and Women.

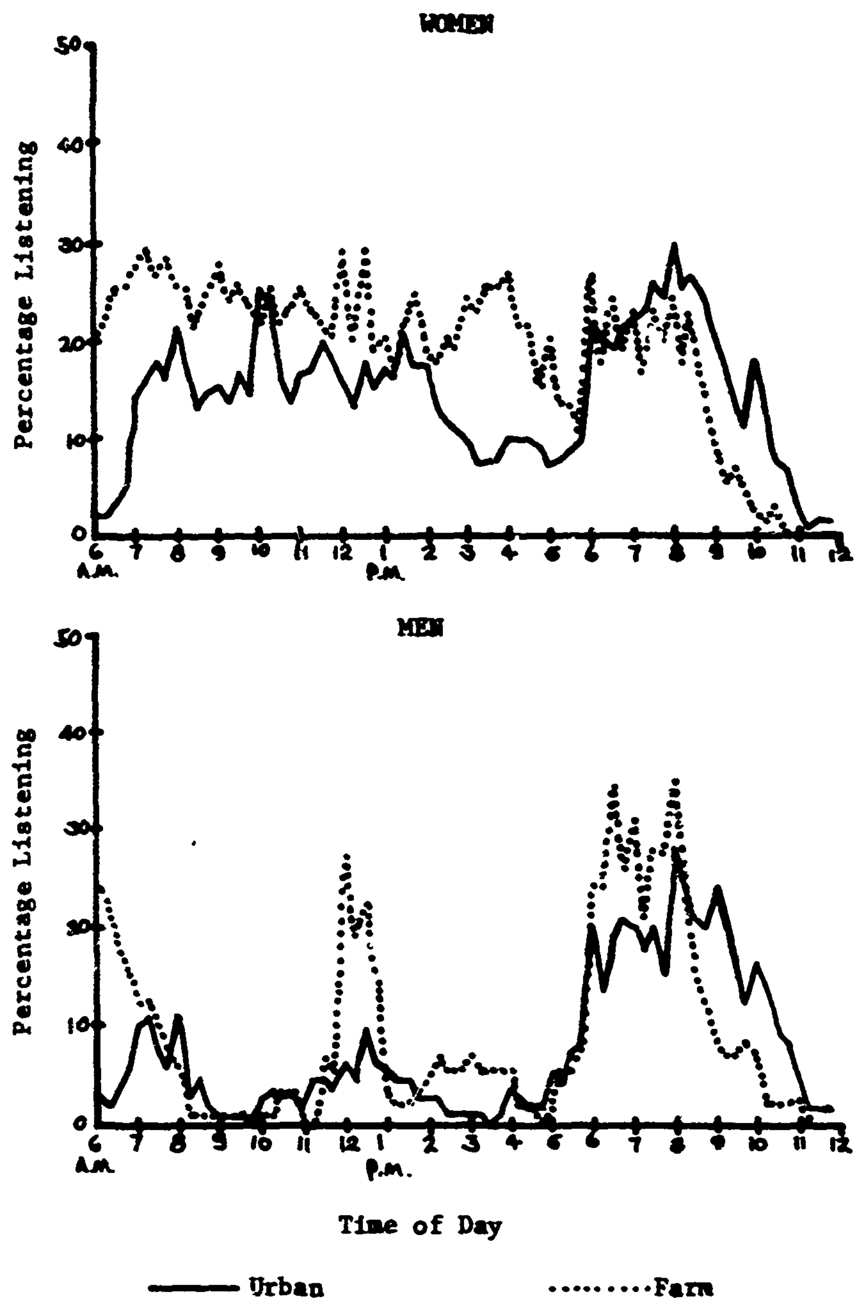


Figure IV-6. Pattern of Listening for Friday, Nov. 8, 1946, by Champaign County Men and Women.

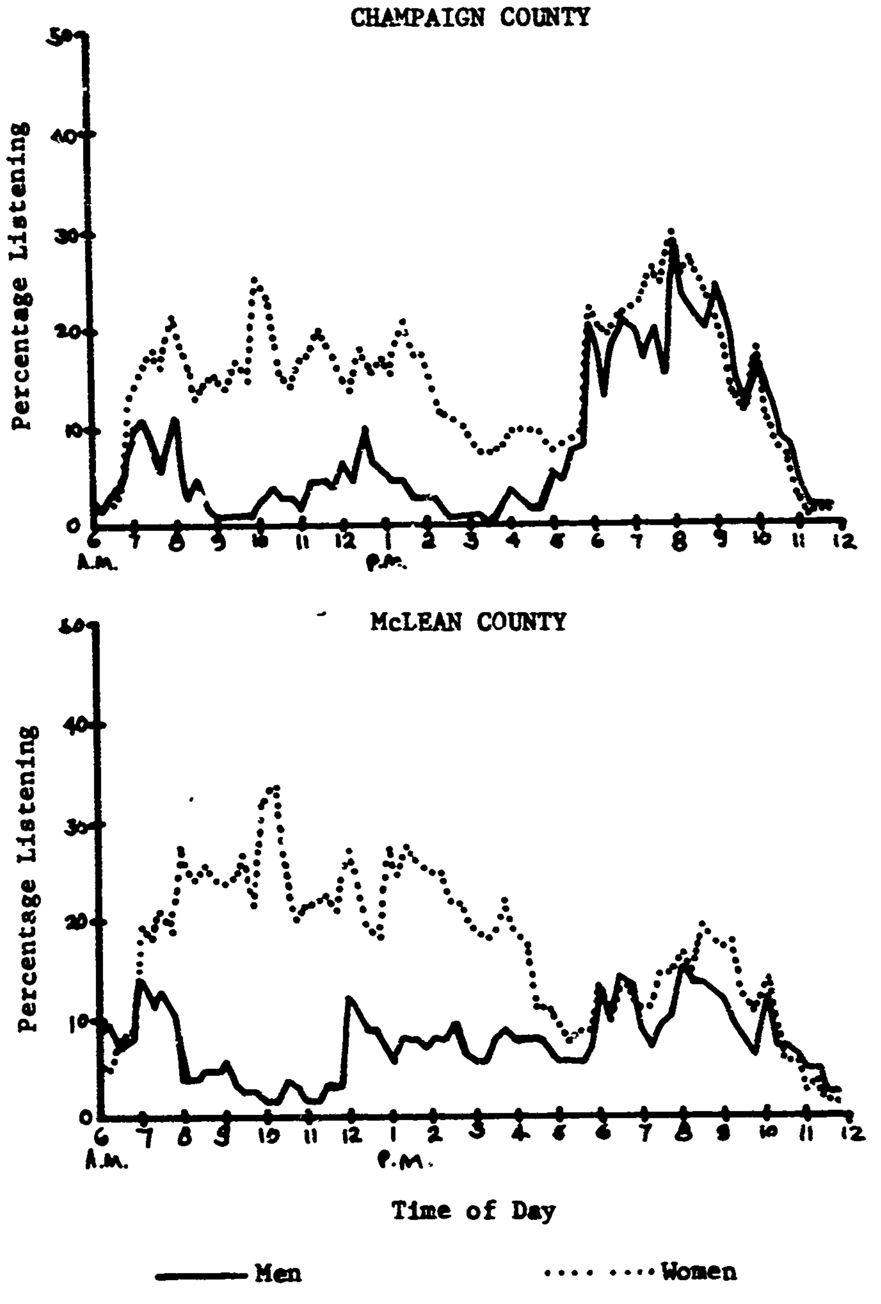


Figure IV-7. Pattern of Listening for Friday, Nov. 8, 1946, by Urban People in Both Counties.

the Hooper "sets-in-use" rating and comparable Nielsen Radio Index ratings for two periods.[12]

The graphs in Figs. IV-3-7 above display not the ratings, but the actual percentages of men and women listening, found by studies in two central Illinois counties in 1946 and 1947. These results are for urban residents, for a weekday and a Sunday, in summer and fall.[13] A rough comparison of Sunday and weekday listening seems to support these assumptions: (1) that early Sunday listening is much like weekday listening except that it happens about

---

[12]The Hooper ratings (Hooper, City Hooperatings) are weekday averages for October, 1947-February, 1948, inclusive. The Nielsen ratings, cited in Sandage, Radio Advertising, p. 133 are weekday averages for the period December 27, 1942-January 23, 1943 and the period July 11-August 7, 1943. It should be noted that the Hooper and Nielsen measurements differ somewhat, Hooper measuring average listening and Nielsen measuring total listening. Except possibly at the extreme ends of the curves, Hooper probably measures 10-30 per cent less than the equivalent Nielsen rating. See the discussion in the "Appendix on Radio Measurement," below.

[13]Results of the two surveys are found in Charles H. Sandage, "Qualitative Analysis of Radio Listening In Two Central Illinois Counties," Bureau of Economic and Business Research Bulletin Series, No. 68, and University of Illinois Bulletin, Vol. 46, No. 50 (March, 1949). The diary method was used to record listening. Every adult (18 years of age or older) member of a participating family recorded his listening by 15 minute periods from 6:00 A.M. to midnight for seven consecutive days. The families were a cross section of families in two central Illinois counties and provided a sample of 790 adults in McLean County and 528 adults in Champaign County. Each of these counties contains an urban center of about 40,000 permanent residents, several villages, and an extensive farm population. The Champaign County study covered the week beginning November 3, 1946, McLean County the week beginning July 27, 1947.

one hour later, and (2) that late Sunday listening is much like late listening on other days. Finally, to get the early and late weekday ratings, we have chosen to sketch in a continuation of the Hooper ratings in Figure IV-2 following roughly the Nielsen curves. This gives the following sets-in-use ratings: 6:00 A.M.-2.0, 7:00 A.M.-8.5, 11:00 P.M.-18.0, 12:00 P.M.-5.0, 1:00 A.M.-1.0. Then from the ratings, the appropriate "shares of the audience," the broadcast schedule, and the assumption that Saturday listening is much like weekday listening, we have completed the news broadcast ratings (Table IV-1 below).

The next step is to turn these ratings into audiences. For this purpose, we need data on the number of listeners per listening household, i.e. data on listening densities, by hours of the day and days of the week.

Listening Densities

In addition to the daily and hourly variation in the number of homes listening to radio (in the case of the Hooper ratings, we have equated the number of sets-in-use and the number of households listening), there is also a daily and hourly variation in the number and kind of listeners per household. This situation is illustrated by the graph in Fig. IV-8.[14] The audience and number of

[14] All the data in this section are taken from Chapter 5 of Eugene F. Seehafer and Jack W. Laemmar, Successful Radio and Television Advertising (New York: McGraw-Hill Book Co., Inc., 1951). The graph is found on p. 105. The authors cite CBS as the source of the data; however, no publication or date is provided.

Table IV-1. Average Ratings for Hourly News Broadcasting

Hour	Days of the Week		
	Monday-Friday	Saturday	Sunday
6:00 A.M.	1.6	1.6	0.0
7:00	8.5	8.5	1.2
8:00	14.3	14.0	6.6
9:00	7.2	11.5	4.9
10:00	4.5	11.4	4.1
11:00	14.2	15.3	10.0
12:00	14.6	18.4	22.1
1:00 P.M.	16.7	13.6	16.1
2:00	1.6	3.6	2.3
3:00	5.7	12.3	7.8
4:00	7.4	16.2	2.9
5:00	11.0	17.6	3.7
6:00	32.0	30.0	14.8
7:00	33.3	30.8	2.0
8:00	22.4	13.0	3.0
9:00	7.1	6.0	1.9
10:00	.4	2.2	0.0
11:00	15.5	15.5	15.5
12:00	4.3	3.3	4.3
1:00 A.M.	1.0	1.0	1.0

Notes: These ratings were calculated from Hooper program ratings and vary both with those ratings and with the schedule of news broadcasts. The calculations are described above.

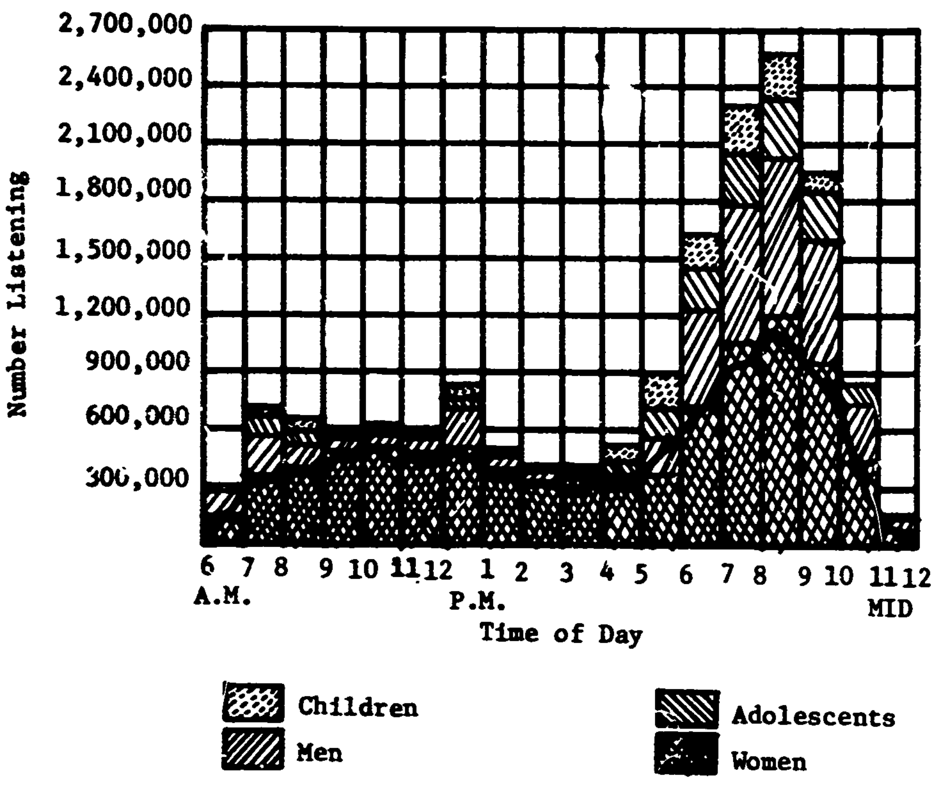


Figure IV-8. Weekday Radio Audience by Hours. The heavy dark line indicates the number of families in the audience during all hours of the day. (CBS.)

listening families of radio station WBBM, Chicago, are depicted for each hour from 6:00 A.M. to 12:00 P.M. for an average weekday. The audience increases slowly through the morning hours, declines somewhat in the afternoon, and then rises sharply in the evening until about 8:30 P.M.; thereafter it declines sharply again.

The composition of the audience also changes. During both the very early morning hours and the evening hours, nearly as many men as women are in the audience; during the work hours, 9:00 A.M. to 5:00 P.M., the audience is composed largely of women. However, at no time of the day are there as many men as women in the audience.[15]

Equating the number of families listening with the number of sets-in-use, we can then make a rough calculation from the graph of the listening densities for men and women for each hour of the day. These are shown in Table IV-2 below and also plotted in Fig. IV-12.

In addition to the data for the single Chicago radio station, Seehafer also cites (p. 106-107) national densities (Tables IV-3 and IV-4 below) from one of the Hooper coast-to-coast studies. The date given is 1949; however, it is not clear whether this refers to the data of the study, or

---

[15]This general pattern was also found in the two diary studies reported by Sandage in Radio Listening in Two Central Illinois Counties. See the graphs in Figs. IV-2 to 7.

Table IV-2. Approximate Listening Densities for Men and Women for Radio Station WBBM

Time	Densities (Listeners per Set-in-Use)	
	Men	Women
6:30 A.M.	.80	1.00
7:30	.59	1.09
8:30	.27	1.04
9:30	.13	1.07
10:30	.14	1.18
11:30	.14	1.10
12:30 P.M.	.45	1.14
1:30	.15	1.18
2:30	.13	1.10
3:30	.19	1.06
4:30	.19	.92
5:30	.39	.75
6:30	.77	1.07
7:30	.76	1.11
8:30	.78	1.13
9:30	.77	1.14
10:30	.66	1.00
11:30	.60	1.08

Table IV-3. Composition of the Radio Audience - Daytime. (Number of listeners per listening set)

New York Time	Audience	Mon.-Fri. Average	Sat.	Sun.	New York Time	Audience	Mon.-Fri. Average	Sat.	Sun.
8-9 A.M.	Women	1.04	1.02		1-2 P.M.	Women	1.08	1.07	1.15
	Men	0.37	0.43			Men	0.28	0.52	0.85
	Children	0.36	0.52			Children	0.23	0.49	0.39
	Total	1.77	1.97			Total	1.59	2.08	2.39
9-10 A.M.	Women	1.08	1.03		2-3 P.M.	Women	1.10	1.12	1.18
	Men	0.27	0.51			Men	0.25	0.51	0.84
	Children	0.32	0.62			Children	0.19	0.30	0.40
	Total	1.67	2.16			Total	1.54	1.93	2.42
10-11 A.M.	Women	1.09	0.99		3-4 P.M.	Women	1.07	1.06	1.12
	Men	0.24	0.38			Men	0.23	0.54	0.75
	Children	0.26	0.67			Children	0.21	0.29	0.37
	Total	1.59	2.04			Total	1.51	1.89	2.24
11-12 A.M.	Women	1.08	0.90		4-5 P.M.	Women	1.08	1.07	1.10
	Men	0.24	0.36			Men	0.25	0.59	0.75
	Children	0.23	0.66			Children	0.31	0.34	0.40
	Total	1.55	1.92			Total	1.64	2.00	2.25
12-1 P.M.	Women	1.08	1.08	1.14	5-6 P.M.	Women	1.01	1.05	1.11
	Men	0.26	0.44	0.76		Men	0.38	0.71	0.83
	Children	0.24	0.44	0.38		Children	0.49	0.40	0.42
	Total	1.58	1.96	2.28		Total	1.88	2.16	2.36

Source: C. E. Hooper, Inc., 1949.

Table IV-4. Composition of the Radio Audience - Evening. (Number of listeners per listening set)

New York Time	Audience	Sun.	Mon.	Tue.	Wed.	Thur.	Fri.	Sat.	Mon.-Fri. Average	Sun.-Sat. Average
6-7 P.M.	Women	1.09	1.00	1.00	1.04	1.06	1.02	1.08	1.024	1.04
	Men	0.85	0.60	0.61	0.65	0.61	0.61	0.74	.616	0.67
	Children	0.50	0.48	0.50	0.50	0.47	0.48	0.43		0.48
	Total	2.44	2.08	2.11	2.19	2.14	2.11	2.25		2.19
7-8 P.M.	Women	1.21	1.05	1.09	1.03	1.06	1.03	1.10	1.052	1.09
	Men	0.95	0.72	0.75	0.73	0.72	0.72	0.80	.728	0.78
	Children	0.52	0.54	0.49	0.60	0.52	0.53	0.45		0.52
	Total	2.68	2.31	2.33	2.36	2.30	2.28	2.35		2.39
8-9 P.M.	Women	1.26	1.06	1.05	1.07	1.05	1.10	1.15	1.066	1.11
	Men	0.96	0.74	0.74	0.75	0.74	0.77	0.80	.748	0.79
	Children	0.45	0.45	0.47	0.50	0.48	0.50	0.48		0.47
	Total	2.67	2.25	2.26	2.32	2.27	2.37	2.43		2.37
9-10 P.M.	Women	1.17	1.12	1.13	1.12	1.13	1.13	1.11	1.126	1.13
	Men	0.88	0.75	0.80	0.79	0.75	0.80	0.84	.778	0.80
	Children	0.33	0.37	0.36	0.34	0.38	0.41	0.44		0.37
	Total	2.38	2.24	2.29	2.25	2.26	2.34	2.39		2.30
10-10:30 P.M.	Women	1.17	1.10	1.13	1.09	1.07	1.04	1.12	1.086	1.10
	Men	0.86	0.77	0.82	0.78	0.75	0.80	0.86	.784	0.80
	Children	0.30	0.35	0.29	0.28	0.30	0.36	0.36		0.32
	Total	2.33	2.22	2.24	2.15	2.12	2.2[illegible]	2.34		2.22

Source: C. E. Hooper, Inc., 1949.

of its publication, or both. These data probably represent a sample of the non-toll call telephone homes of the thirty or so largest U.S. cities sometime during 1949.[16] Since these are the only listening density data available, we shall use them in the audience estimation; however, we might ask several questions about the data.

It appears likely that the data were collected at least a year after the study period. This raises the possibility that television listening might have begun to change the patterns of radio listening by mid-1949. We can offer some evidence that that was not the case: first, there were approximately twenty times as many radio as television homes in 1949 (39,280,000 radio homes to 1,960,000 television homes); second, as the table below indicates, the average number of hours listened per day per (Nielson) home changed very little during the period.[17]

---

[16]On pp. 89-90 of his Radio Audience (1944), Hooper describes his sample for such data as "the telephone home population of the thirty-two large cities of equal network opportunity." He does not provide his age break between adults and children.

[17]The radio datum is a Broadcast Measurement Bureau estimate cited in Schaefer, p. 102. The television estimate, also cited in Schaefer (p. 104), is from Broadcasting, July 10, 1950. The table is reproduced from Schaefer, p. 108. The source is the A. C. Nielsen Company.

Average Number of Hours Listened Per Day Per NRI Home
Annual Average, 1946-1949

Year	Hours
1946	4.0
1947	4.3
1948	4.4
1949	4.2

If there were no changes due to the introduction of television, were there possibly seasonal changes in listening densities? It is generally assumed, probably on the basis of ratings, that listening increases in the fall and winter. The ratings, equivalent to the percentage of sets-in-use, do indeed show seasonal fluctuations as the graphs in Figs. IV-9, IV-10, and IV-11 demonstrate. However, strictly speaking, the increase in the ratings from summer to winter does not necessarily imply an increase in the audience; only if the listening density is assumed not to decline very much can we be confident that the audience has increased. The tacit assumption seems to be that the listening density is rather constant, and it is upon this assumption that we proceed. Thus it is consistent with all the data and rather plausible that the listening density was rather constant at least during the period from 1947 to 1949.

The data from these tables have been plotted together in Fig. IV-12.[18] In general, the shapes of the WBBM and

[18] The Hooper data for weekday evenings has been averaged to produce a single weekday (Monday-Friday) evening plot. Also, since the form is hourly averages, the values have been plotted on the half hour, e.g., the value for 8:00-9:00 P.M. has been plotted at 8:30 P.M.

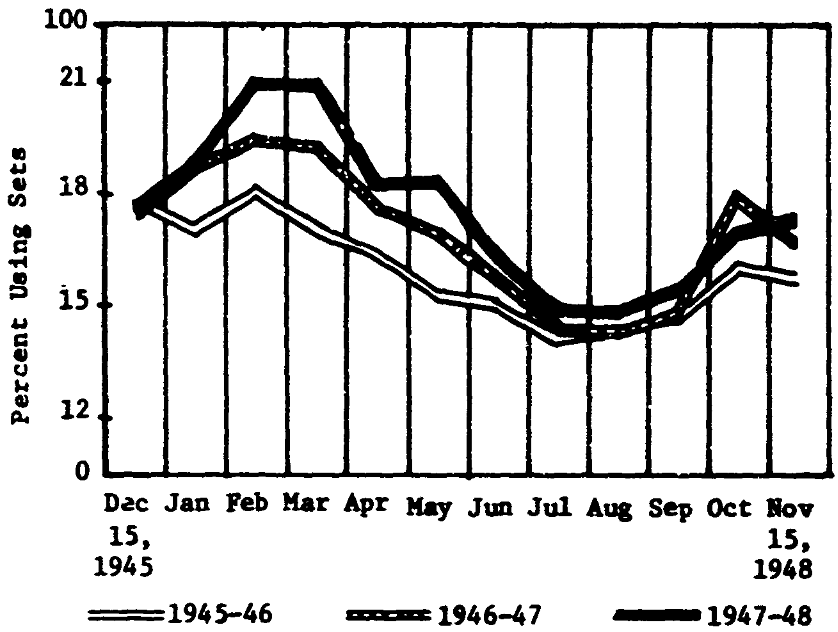


Figure IV-9. Sets-In-Use Index for Daytime (Mon.-Fri., 8:00 A.M.- 6:00 P.M.). (Hooper Basic Audience Trends)

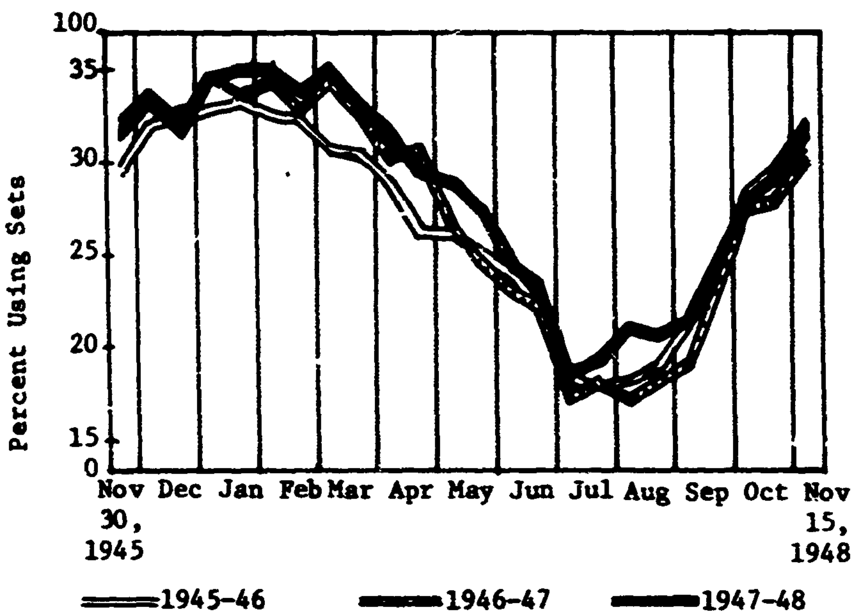


Figure IV-10. Sets-In-Use Index for Evenings (Sun.-Sat., 6:00-10:30 P.M.). (Hooper Basic Audience Trends)

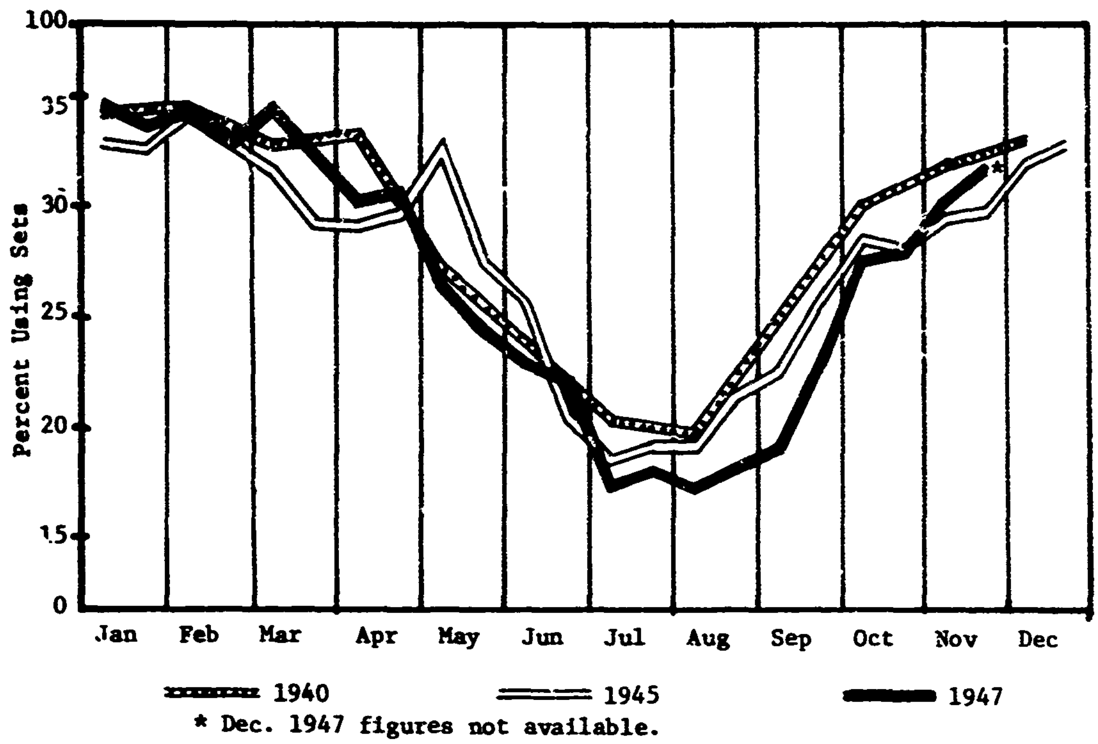


Figure IV-11. Sets-In-Use Index for Evenings (Sun.-Sat., 6:00-10:30 P.M.). (Hooper Basic Audience Trends)

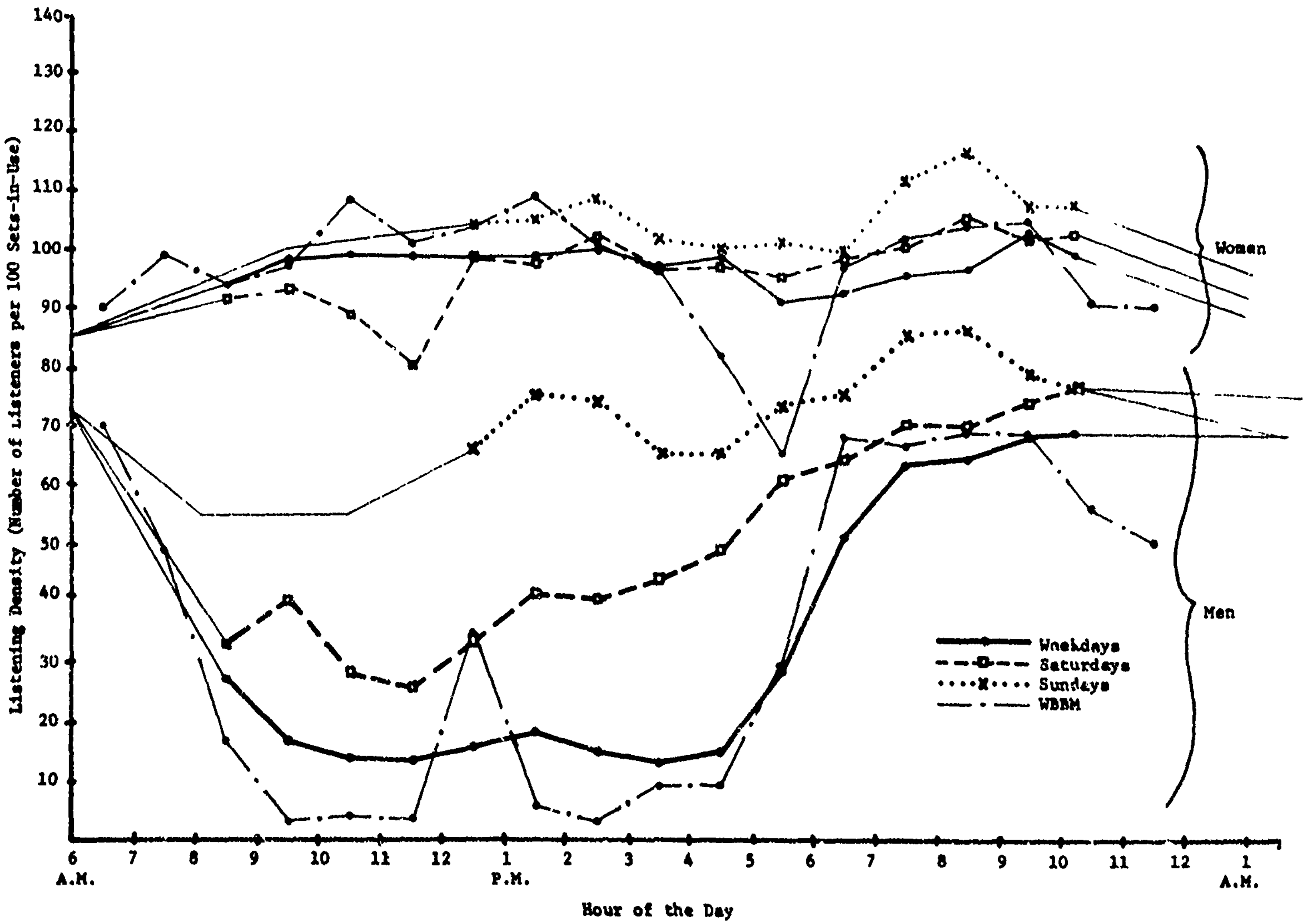


Figure IV-12. Graph of Hourly Listening Densities for Men and Women (Weekdays, Saturdays, and Sundays).

Hooper curves for men and women are quite similar; the two obvious discrepancies, for men at 12:30 A.M. and for women at 5:30 P.M., may be due to peculiarities of programming by the single station. At any rate, since the Hooper curves are more accurate, and since the Hooper sample consists of residents of cities, we have in general relied on the Hooper data.

However, the Hooper data are incomplete, lacking the listening densities for the Sunday morning hours and for the early morning and late night hours for each day of the week. We have completed the missing curves, generally paralleling the shapes of the known WBBM and Hooper data. The reasoning is as follows: both the WBBM and the Sandage data indicate that although the audience is relatively smaller during very early morning and very late evening hours, there is a nearly equal number of men and women at those times. Therefore the listening densities must be nearly equal.[19] This means, in general, that the listening density for women must decline at the extremes of the day, and that for men must increase. The rate of the respective changes was judged from the WBBM curve. Since the densities for men at 10:30 P.M. were approximately equal to those at 6:30 A.M., we have left them nearly constant

---

[19] This is obvious, since the listening density for each sex is just the number listening of that sex divided by the total numbers of sets in use.

from 10:30 P.M. to 1:00 A.M. Thus have all the listening densities been computed. The data and method here are obviously inexact; however, we believe that they are as precise as warranted, given the inaccuracies in the content analysis and message exposure probabilities, which are dealt with later.[20] The final values of the densities are presented in Tables IV-6, IV-7, and IV-8 below.

The News Broadcast Audiences

We now calculate the audiences for news broadcasts. The ratings are approximately the percentages of households listening. The densities are the number of male and female adult listeners per hundred households. The product of the ratings and the total number of households gives the number of listening households. This number multiplied by the densities gives the number of adult male and female listeners.[21]

---

[20]Also, since the audiences are calculated as the <u>product</u> of the ratings and densities, and since the ratings are <u>very</u> small in the very early morning and very late evening <u>hours</u>, the large percentage errors in the densities should not significantly change the overall daily audience structure.

[21]Actually, not all adult males and females lived in households. According to the 1950 census of the Cincinnati Standard Metropolitan Area, within the 276,715 households of the area lived 874,510 of the 904,402 total population. The institutional population was 9,297, leaving 20,595 members of the population (904,402 minus 874,510 minus 9,297) "at large" but not living in households. We have ignored this discrepancy in making the calculations.

Listening Outside the Home

The Hooper data, upon which the audience sizes have been based, measure only the listening in the home. About 1949, as it was becoming obvious that television was going to diminish the size of the radio audience, several studies were conducted to measure the size of the audience outside the home. These studies found that there were indeed a significant number of outside listeners, often numbering as much as 30 or 40 percent of the audience at home.[22] There are no Cincinnati or national data among these studies; therefore we have had to adapt them for the Cincinnati audience.

Table IV-5 below shows Pulse data on the hourly in-home and out-of-home listening in the New York City area during August 1949.[23]

Graphs of out-of-home listeners as a percent of in-home listeners for weekdays, Saturday, and Sunday, are shown in Fig. IV-13. Also, this percentage has been extrapolated for each of three curves to provide values for the early morning and late evening hours.

---

[22]Several of these studies are cited in H. M. Beville, Jr., The True Dimensions of the Radio and Television Audience (New York: National Broadcasting Co., Inc., 1949).

[23]Cited by Beville, True Dimension, p. 25.

Table IV-5. Comparison of In-Home and Out-of-Home Listening

Time	Listeners Out-of Home (in thous.)	Listeners in Home (in thous.)	Outside Listeners As Percent of Inside Listeners
Monday-Friday			
9-10 A.M.	134.4	769.9	17.5%
10-11	168.0	1,165.5	14.4
11-12	134.4	1,150.2	11.7
12-1 P.M.	190.4	1,028.1	18.5
1-2	246.4	838.2	29.4
2-3	436.9	946.4	46.2
3-4	616.1	1,051.6	58.6
4-5	369.7	1,171.8	31.5
5-6	190.4	1,318.3	14.4
6-7	156.8	1,649.7	9.5
7-8	67.2	1,850.0	3.6
8-9	123.2	1,810.0	6.8
Daily Avg.	236.2	1,228.5	19.2%
Saturday			
9-10 A.M.	33.6	555.9	6.0%
10-11	145.6	1,197.2	12.1
11-12	89.6	1,379.6	6.5
12-1 P.M.	156.8	1,143.6	13.7
1-2	168.0	975.2	17.2
2-3	369.7	1,215.9	30.4
3-4	347.3	1,267.6	27.4
4-5	358.5	1,391.5	25.8
5-6	235.2	1,367.7	17.2
6-7	212.8	1,521.7	14.0
7-8	134.4	1,489.8	9.0
8-9	358.5	1,320.1	27.2
Daily Avg.	217.5	1,235.5	17.6%

Table IV-5 (Continued)

Time	Listeners Out-of Home (in thous.)	Listeners in Home (in thous.)	Outside Listeners As Percent of Inside Listeners
Sunday			
9-10 A.M.	33.6	776.2	4.3%
10-11	89.6	1,109.8	8.1
11-12	112.0	1,170.5	9.6
12-1 P.M.	78.4	1,283.2	6.1
1-2	201.6	1,100.1	18.3
2-3	369.7	1,439.1	25.7
3-4	481.7	1,742.	27.7
4-5	616.1	2,065.7	29.8
5-6	627.3	1,984.3	31.6
6-7	403.3	2,197.1	18.4
7-8	795.3	2,182.7	36.4
8-9	997.0	2,397.1	41.6
Daily Avg.	400.5	1,620.7	24.3%

Source: The Pulse, "Radio Listening Out of Home in New York," August, 1949.

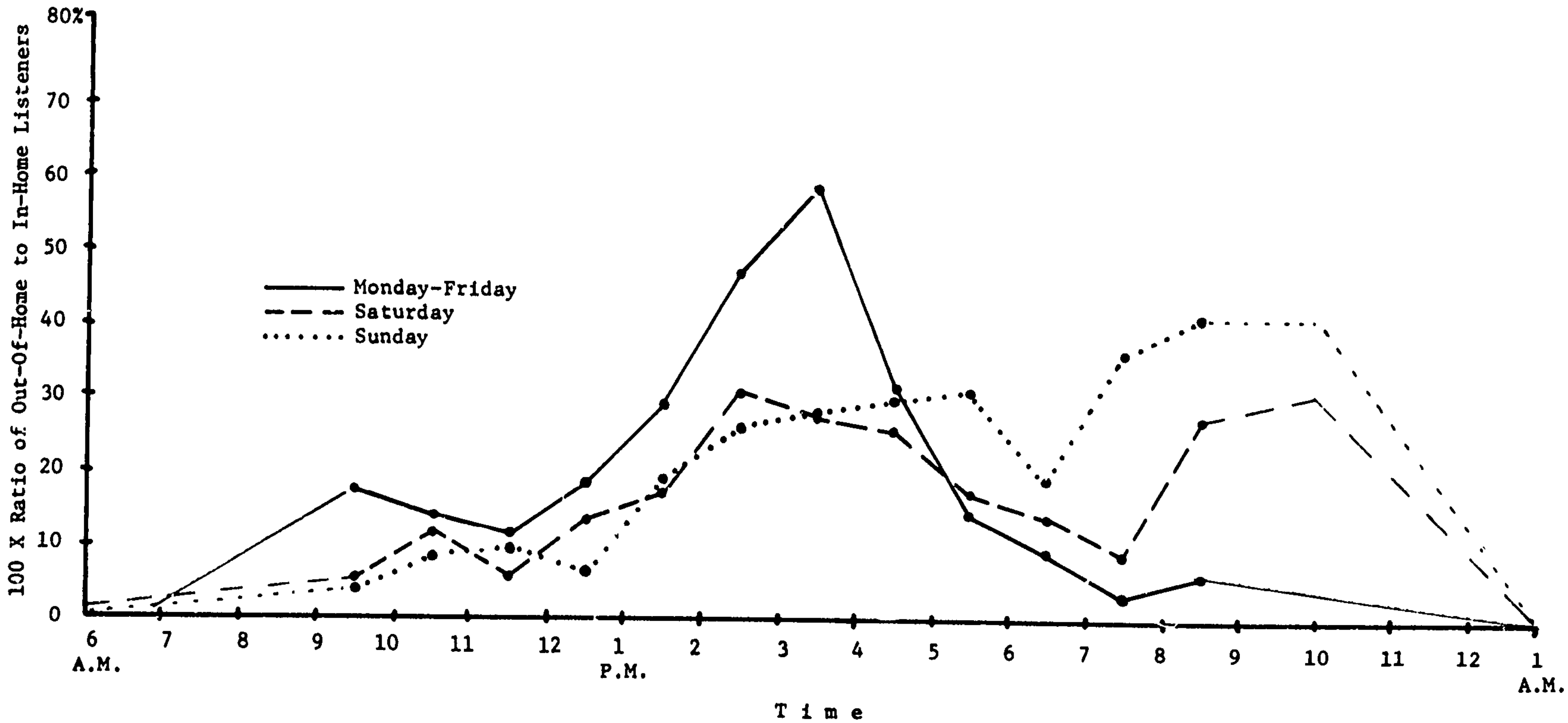


Figure IV-13. Out-Of-Home Listeners as a Percent of In-Home Listeners for Weekdays, Saturdays, and Sundays.

In addition to the data above, Beville also cites some data on out-of-home listening by sex. The Psychological Corporation, in a study of the Des Moines, Iowa, and Springfield, Massachusetts, areas, found (not surprisingly) that males do a much larger proportion of their day's listening outside the home, than do females. This also holds true for single people compared with married people. The data are presented below.[24]

	Males		Females		Total Group
	Single	Married	Single	Married	
<u>Average Number of Minutes of Listening</u>					
Outside Listening	56	30	19	17	27
Inside Listening	111	161	192	258	196
Total	167	191	211	275	223
<u>Per Cent of Total Day's Listening</u>					
Outside Listening	34%	16%	9%	6%	12%
Inside Listening	66	84	91	94	88
Total	100%	100%	100%	100%	100%

We have used these data to adjust the radio audiences to include outside listening. For each hour of the

[24]The data are cited by Beville, <u>True Dimensions</u>, p. 25.

day, the total audience has been increased by the proportion given in the graphs (Fig. IV-13) of the Pulse data. However, since the Psychological Corporation data show that the per cent of the total day's outside listening for males is about three times that for females (the married males: 16 per cent; married females: 6 percent; single males: 34 per cent; single females: 9 percent), we have required the proportional increase for males to be three times that for females.[25] This then allows the correction to account for outside listening for each hour of the day. The final values of the audiences are presented in Tables IV-6, IV-7, and IV-8 below.

---

[25] The equations are quite simple. Let:

L_m = the number of male in-home listeners at any time.
L_f = the number of female in-home listeners at any time.
P_m = the proportion by which the male in-home listening is to be increased.
P_f = the proportion by which the female in-home listening is to be increased.
P = the total proportion by which the in-home listening is to be multiplied to account for outside listening.

Since the total increase must equal the sum of the increase in males and the increase in females we have:

$$P_m \cdot L_m + P_f \cdot L_f = P \cdot (L_m + L_f) .$$

Also, the proportional increase for males is to be three times that for females, i.e.

$$P_m = 3P_f$$

These equations are then solved for P_m and P_f at each hour to give the proportion by which the original male and female listenership is to be increased to account for out-of-home listening.

Table IV-6. News Broadcast Ratings and Audiences, and Listening Densities, Monday - Friday, 6:00 AM - 1:00 AM

Time	News Broadcast Ratings	Listening Densities (Number of Listeners per Hundred Listening Homes)		Audiences			
		Men	Women	Number of Men	Percentage	Number of Women	Percentage
6:00 A.M.	1.6	82.5	95.0	3,800	1.31	4,200	1.28
7:00	8.5	64.5	98.0	15,800	5.44	23,400	7.16
8:00	14.3	46.0	102.0	20,900	7.19	42,400	12.97
9:00	7.2	32.0	106.0	8,300	2.86	23,300	7.12
10:00	4.5	25.5	108.5	4,300	1.48	15,100	4.62
11:00	14.2	24.0	108.5	12,200	4.20	46,700	14.28
12:00	14.6	25.0	108.0	13,400	4.61	48,400	14.80
1:00 P.M.	16.7	27.0	108.0	18,600	6.40	58,100	17.77
2:00	1.6	26.5	109.0	2,100	.72	6,100	1.87
3:00	5.7	24.0	108.5	7,600	2.62	23,700	7.25
4:00	7.4	24.0	107.5	9,800	3.37	29,300	8.96
5:00	11.0	31.5	104.5	14,100	4.85	36,800	11.25
6:00	32.0	49.8	101.7	53,700	18.48	96,600	29.54
7:00	33.3	67.4	103.8	69,000	23.75	99,200	30.33
8:00	22.4	73.8	105.9	49,700	17.11	67,500	20.64
9:00	7.1	76.3	109.6	16,500	5.68	22,200	6.79
10:00	.4	78.0	110.0	900	.31	1,200	.37
11:00	15.5	78.0	106.0	34,800	11.98	46,100	14.10
12:00	4.3	78.0	102.0	9,500	3.27	12,200	3.73
1:00 A.M.	1.0	78.0	98.0	2,200	.76	2,700	.83

Table IV-7. News Broadcast Ratings and Audiences, and Listening Densities, Saturday 6:00 AM - 1:00 AM

Time	News Broadcast Rating	Listening Densities (Number of Listeners per Hundred Listening Homes		Audiences			
		Men	Women	Men		Women	
6:00 A.M.	1.6	82.5	95.0	3,800	1.31%	4,200	1.28%
7:00	8.5	67.0	97.5	16,500	5.68	23,300	7.12
8:00	14.0	51.5	100.0	21,300	7.33	39,600	12.11
9:00	11.5	47.0	102.5	16,300	5.61	33,600	10.27
10:00	11.4	44.5	101.0	16,400	5.65	33,600	10.27
11:00	15.3	37.0	94.5	18,500	6.37	42,400	12.97
12:00	13.4	40.0	99.0	24,300	8.36	53,600	16.39
1:00 P.M.	13.6	48.0	107.5	23,400	8.05	44,400	13.58
2:00	3.6	51.5	109.5	7,400	2.55	12,5[illegible]	3.82
3:00	12.3	52.5	109.0	27,300	9.40	43,600	13.33
4:00	16.2	56.5	106.5	37,300	12.84	55,200	16.88
5:00	17.6	65.0	106.0	43,300	14.90	57,900	17.70
6:00	30.0	72.5	106.5	75,700	26.06	96,000	29.36
7:00	30.8	77.0	109.0	78,000	26.85	98,700	30.18
8:00	13.0	80.0	112.5	37,300	12.84	44,500	13.61
9:00	0.0	82.0	113.5	0	0	0	0
10:00	2.2	85.0	111.5	7,800	2.68	7,900	2.42
11:00	15.5	86.0	109.0	48,600	16.73	51,700	15.81
12:00	3.3	85.0	105.0	9,000	3.10	10,100	3.09
1:00 A.M.	1.0	84.5	102.0	2,300	.79	2,800	.86

Table IV-8. News Broadcast Ratings and Audiences, and Listening Densities, Sunday, 6:00 AM - 1:00 AM

Time	News Broadcast Rating	Listening Densities (Number of Listeners per Hundred Listening Homes		Audiences			
		Men	Women	Men		Women	
6:00 A.M.	0.0	82.5	95.0	0	0.00%	0	0.00%
7:00	1.2	74.0	99.5	2,500	.86	3,300	1.01
8:00	6.6	66.0	103.5	12,600	4.34	19,200	5.87
9:00	4.9	65.0	108.0	9,300	3.20	14,900	4.56
10:00	4.1	65.0	111.0	8,200	2.82	13,000	3.98
11:00	10.0	67.5	112.0	21,500	7.40	32,600	9.97
12:00	22.1	73.0	113.5	50,500	17.38	72,400	22.14
1:00 P.M.	16.1	80.5	114.5	43,100	14.84	54,400	16.63
2:00	2.3	84.5	116.5	7,300	2.51	8,300	2.54
3:00	7.8	79.5	115.0	24,700	8.50	28,500	8.71
4:00	2.9	75.0	111.0	8,900	3.06	10,300	3.15
5:00	3.7	79.0	110.5	12,200	4.20	13,200	4.04
6:00	14.8	84.0	110.0	48,200	16.59	51,100	15.63
7:00	2.9	90.0	115.0	7,200	2.48	7,300	2.23
8:00	3.0	95.5	123.5	12,900	4.44	12,400	3.79
9:00	1.9	92.0	121.5	8,000	2.75	7,800	2.39
10:00	0.0	86.0	116.0	0	0.00	0	0.00
11:00	15.5	84.5	113.5	52,100	17.93	55,800	17.06
12:00	4.3	82.0	109.0	11,800	4.06	13,900	4.25
1:00 A.M.	1.0	79.0	105.5	2,200	.76	2,900	.89

Radio Listening by Age and Education

We have now calculated the total audience and its breakdown by sex for each of the radio vehicles. A pattern quite prominent in these data shows women always outnumbering men in the audience. Because of the division of labor by sex in the culture, women had much more opportunity (and possibly greater need) to listen. Even during the evening hours when the opportunity to listen was relatively more equal, the women seemed to be more inclined to listen.

This interaction between listening and sex is the most important which we have found. In the appendix of their autumn, 1947 radio study, Lazarsfeld and Kendall report on the relationship between listening and several other of our population dimensions.[26] Table IV-9 shows only a slight association between amount of evening listening and age, i.e., younger people seem to listen somewhat more. This tendency is also found in Table IV-10 which shows the proportion of heavy listeners by age and education.

However, we find a stronger negative association between education and heavy listening: at every age level, those with college education seem to listen less. Insofar as listening to our "news-on-the-hour" vehicles is explained

---

[26]The three tables which follow are taken from Appendix C, pp. 131-146 of Paul A. Lazarsfeld and Patricia L. Kendall, Radio Listening in America: the People Look at Radio--Again (New York: Prentice-Hall, Inc., 1948).

Table IV-9. Amount of Evening Listening by Age

Amount of Evening Listening	Age		
	21-29	30-40	50-
Less than one hour	20%	24%	27%
1-3 hours	48	49	45
3 or more hours	32	27	28
	100%	100%	100%

Source: See footnote 26.

Table IV-10. Proportion of Heavy Listeners[a] by Age and Education

	Age		
Education	21-29	30-49	50-
College	18%	20%	17%
High School	36	28	28
Grade School	33	29	31

[a]Heavy listeners are those who listen to the radio three or more hours on an average weekday evening.

Source: See footnote 26.

more by total listening time rather than selective tuning, this negative association between education and amount of listening would imply less news exposure for the college educated. However, the better educated do show a decided preference for both news broadcasts and public affairs programs. Table IV-11 shows that at each age level, those with higher education are more likely to prefer news broadcasts. Despite the lesser amount of listening by the college educated, they may, because of selective tuning, be as much or more exposed to news broadcasting as those of lesser education.

Table IV-11. Proportion Choosing News Broadcasts Among Evening Program Preferences by Age and Education

	Age		
Education	21-29	30-49	50-
College	75%	78%	83%
High School	72	75	79
Grade School	61	70	74

Source: See footnote 26.

Thus, the younger and less well educated listen more, but the older and better educated are more likely to prefer news broadcasts. On the basis of this evidence and the lack of data specifying additional audience breakdowns

for the Cincinnati news vehicles, we have used only the tables of audience by sex in creating the simulation radio vehicles.

Other Radio Information

For the purposes of this simulation, we will assume that the news broadcast vehicles defined above carried all the significant international and United Nations radio news. Other news, in the form of special programming, e.g., information programs and spot announcements, was broadcast over individual stations and therefore would require defining additional vehicles if it were to be included in the simulation.

During the course of the educational campaign, several special United Nations information programs were presented. Mr. Robert Adair, the public relations director for the Cincinnati Plan, recalls that there were four or five of these programs, probably broadcast on Sunday afternoons.[27] Since we have no additional data about the content of these programs, and since they account for a very small proportion of the total radio information, we have not defined a special vehicle to include them in the simulation.

The second known omission from the radio broadcast information are "spot" announcements about the United Nations.

---

[27]In conversation with the author.

Information about these "spots" is sparse, consisting only of passing references in the final report on the educational campaign and in an article on the campaign in The New York Times Magazine. From the report:

> . . . The radio stations broadcast facts about the United Nations, one of them scheduling spot programs one-hundred-and-fifty times a week. . . .
>
> . . . The slogan, "Peace Begins with the United Nations--the United Nations Begins with You," which had been broadcast at the end of one-minute "spot" announcements about the organization one-hundred-and-fifty times a week, was not recalled by fifty-one per cent of the people. . . .

From the magazine article:

> . . . A hundred and fifty times a week through January and February WLW dropped in "spots" between its most popular radio programs: "What is the General Assembly? *** It is the Town Hall of the World." . . .[28]

In order to model these spot announcements, we would have to define an additional twenty-one vehicles, the between-program audiences of station WLW. In view of the effort required and the evident limited amount of information contained by these spots,[29] we have not included them in the simulation scenarios.

---

[28]The two quotations from the report, Star and Hughes, The Cincinnati Plan, are found on pp. 2 and 9. The third quotation is from Gilbert Bailey, "To Make Us Aware," New York Times Magazine, March 7, 1948, pp. 24-25.

[29]Star and Hughes in The Cincinnati Plan, p. 9, judge that the fifty-one per cent who could not recall the slogan was surely an underestimate:

> . . . Another ten per cent claimed that they had heard it but had no idea when or where. Moreover, hearing does not mean understanding. One woman, questioned on the slogan, said: "Why, yes. I heard it over and over again . . . but I never did find out what it means. . . ."

CHAPTER V

CONSTRUCTING THE MEDIA SYSTEM III: GENERATING VEHICLE EXPOSURE PROBABILITIES FROM CUMULATION DATA

The Importance of Cumulation

In Chapter III we generated a population distributed across the 144 audience types defined by the five dimensions. In Chapters III and IV we defined the media vehicles and distributed the audience of each vehicle across the 144 audience types. The ratio of the vehicle audience to the total number of people in an audience type can be considered a mean probability of exposure to the vehicle for members of the given audience type. The next step is to assign probabilities to each member of the audience type so that this overall mean exposure probability is reproduced. There is, obviously, an infinite number of sets of probabilities which fulfill this requirement; how shall we choose from among these sets?

Consider a hypothetical vehicle which has an average audience of 20 per cent of the population, i.e., a mean probability of exposure of 0.20. To simplify matters, let us consider a population consisting of only ten people. We can show that different sets of

probabilities, all of which imply the same mean probability of 0.20, can result in quite different rates of growth of the total number of people exposed (the cumulative audience), or distributions of frequencies of exposure. In Table V-1 below, we have calculated the audience cumulation through five time periods for three possible sets of probabilities. In the first set each probability is the mean value 0.20; in the second set two probabilities have the value 1.00 and the remaining eight are 0.00; in the final set, two probabilities have values equal to 0.60 and the other eight have values of 0.10. We observe from the table that each of these models has an average audience of two people; however, the audience cumulation is quite different. The model for which all the probabilities are equal to the mean probability has a high rate of cumulation; the second model has no cumulation whatsoever--the same two individuals are exposed at each issue; and the third model has a moderate rate of cumulation lying between the other two. Thus, the distribution of probabilities over the population governs the kind of exposure cumulation (and frequencies) produced by the model.

In Appendix B we have outlined the mathematics relating a particular model of the population probabilities to observed (or estimated) cumulation and frequency data. The model divides the population into three groups--the

Table V-1. Growth of the Cumulative Audience in a Population of Ten Individuals Through Five Time Periods for Three Different Probability Distributions.

	Cumulative Number of Individuals Exposed At Least Once		
Time Period	Ten 0.20 Probabilities	Eight 0.0 and two 1.0 Probabilities	Eight 0.10 and two 0.60 Probabilities
1	2.00	2.00	2.00
2	3.60	2.00	3.20
3	4.88	2.00	4.04
4	5.90	2.00	4.70
5	6.72	2.00	5.26

regular users or subscribers of the vehicle, the moderate or casual users of the vehicle, and the very infrequent users of the vehicle. For each of these groups, a beta-function[1] distribution of probabilities is generated from average audience and two-period cumulation data for that group. In addition we need to know the size of each of the three groups.

Thus, for each vehicle, we define three distributions of probabilities. For each distribution (i), we need to estimate the size (K^i), average audience ($C_1{}^i$), and two-period cumulation ($C_2{}^i$). However, there are

[1]For a complete discussion of the model and the definition and graph of the beta-function, see Appendix B.

several constraints upon these quantities. The three distribution sizes must sum to the total population, the three average audiences to the vehicle audience, and the three two-period cumulations to the vehicle two-period cumulation. In addition a very important constraint, arising from the beta function model, relates the average audience and two-period cumulation of each distribution: the proportion of the distribution population in the two-period cumulation may not exceed the difference between twice the audience proportion and the square of the audience proportion. For example, if the average audience is 20 per cent of the population, the maximum value of the two-period cumulation is 36 per cent of the population $[2(.20) - (.20)^2]$. This relation is expressed algebraically as:

$$\left(\frac{C_2}{K}\right) \leq 2 \left(\frac{C_1}{K}\right) - \left(\frac{C_1}{K}\right)^2$$

If we let P_1 represent the audience proportion and P_2 represent the cumulation proportion then the equation above becomes

$$P_2 \leq 2P_1 - P_1^2.$$

A useful concept is the <u>relative</u> two-period cumulation, defined as the ratio of C_2 to C_1 (or P_2 to P_1). From the equation above the relative accumulation is also limited by the average audience:

Relative two-period accumulation = $P_2/P_1 \leq 2 - P_1$.

Since the possible values of P_1 lie between 0.00 and 1.00, the ratio P_2/P_1 must lie between 1.00 and 2.00; moreover, the larger the value of P_1, the smaller must be P_2/P_1. In other words, as the average audience increases, the number not yet exposed (who contribute to the increase of C_2 over C_1) grows smaller. Thus the maximum possible value of the ratio C_2/C_1 decreases. In the limit, as the average audience includes the entire population, the two-period cumulation must just equal the average audience and the ratio C_2/C_1 is identically 1.00.

The point where the maximum cumulation is attained is usually just the case of greatest randomness (or least structure) in the population.[2] At this point, all members of the population have identical probabilities, equal to the proportion of the population in the average audience. Generally, the more we can divide the population into subgroups (such as very frequent listeners versus non-listeners) characterized by probabilities divergent from the overall mean probability of exposure--that is, the more <u>structure</u> we can discover in the audience patterns, the lower the resulting cumulation.

Cumulation occurs when a person not previously

---

[2]This is also the point where the variance of the <u>average</u> audience, cumulations, etc., is the greatest.

exposed to the vehicle, becomes exposed to it. Therefore if each issue of a vehicle tends to expose the same people repeatedly, the vehicle will tend to have a lower relative accumulation. Thus we may expect that those vehicles whose average audience consists in large part of persons with a high probability of exposure should have a low relative accumulation. An example of this might be a magazine with a high subscription ratio or a radio soap opera whose audience likely excludes nearly all of the males and one-half to two-thirds of the females of the population.

Below we have summarized these observations about accumulation for this model:

1) The minimum value of the two-period relative accumulation is 1.0. Its maximum value is $2.0-P_1$ where P_1 is the audience proportion.

2) It appears likely both from the model and from the data that *ceteris paribus* the larger the audience proportion, the smaller the relative accumulation; thus the audience proportion *which has already been calculated for each vehicle* acts as a strong constraint on the other variables.

3) The larger the proportion of the population

in the high probability distribution, the smaller the relative accumulation.

We turn now to the actual calculation of K, C_1, and C_2 for each distribution of each vehicle.

Distributions and Cumulation for Newspapers

We first estimate the proportion of the population in each subscriber (high probability) distribution. Table V-2 below, taken from the ABC Audit Report for the Post shows the distribution of copies by edition and region. From conversation with the Post circulation managers, we have estimated the following distribution pattern for each of the editions:

Edition	Distribution	Percent in City Zone	Of Those in City Zone, Per Cent to Subscribers
1st	all street sales	83%	0%
2nd	mostly suburban areas	7	0
3rd	Kentucky: 1/4 to streets, 3/4 to homes	82	75
4th	mostly city homes	100	100
5th	1/2 to city homes, 1/2 to streets	98	50
6th	all street sales	98	0
Predate	mostly outside city	1	0

Table V-2. Net Press Run, Time, and Distribution of Editions of the *Cincinnati Post*; Thursday, March 18, 1948.

Edition	Press Time	Date	Issue Dated	Net Press Run[a]	Sales Release Location	Approximate Distribution City Zone	Retail Zone	All Other
1st	10:30 a.m.	3/18	3/18	9,265	A-C[b]	83%	3%	14%
2nd	12:15 p.m.	3/18	3/18	15,550	A-C	7	87	6
3rd	12:30 p.m.	3/18	3/18	38,407	A-C	82	11	7[d]
4th	2:05 p.m.	3/18	3/18	33,021	C[c]	100	0	0
5th	3:40 p.m.	3/18	3/18	48,545	A-C	98	1	1
6th	4:30 p.m.	3/18	3/18	14,047	A-C	98	1	1
Predate	4:45 p.m.	3/18	3/19	5,381	A-C	1	12	87

Source: Audit Bureau of Circulations, *Cincinnati Post Audit Report: Twelve Months Ending March 31, 1948* (Chicago: Audit Bureau of Circulations, 1948), p. 2.

Notes: [a]These figures include copies spoiled in distribution, free copies, unsold and allowances.

[b]A - immediate sales release in City; C - sales release on arrival at destination in Retail Zone and all other.

[c]Late news, home edition.

[d]Distribution in Kentucky area.

The product of the size of each edition, the per cent distributed in the city zone, and the proportion of city zone copies going to subscribers, gives the number of copies going to subscribers. This number is approximately 80,000 copies or 64 per cent of city zone sales. We have used this proportion for each of the daily newspapers. If we assume that this is also the number of subscriber households and use the previously calculated estimate of 2.23 adults per household (see p.110 above) we can calculate the number of persons in the subscriber distributions for each paper:

Calculation of the Number of Newspaper Subscribers

	Newspapers		
	Post	Times-Star	Daily Enquirer
City Zone Circulation	125,008	132,441	103,500
@64% into subscriber households	80,000 homes	84,700 homes	66,300 homes
@2.23 adults per household	178,000 adults	189,000 adults	148,000 adults
% of all adults in subscriber distribution	29%	31%	24%

Since the number of adult readers per copy is about 2.29[3] but the number of adults per household is at most

[3]This was calculated above from the corrected NORC survey data.

2.23, the mean exposure probability for each of these adults must be rather high. We have chosen a mean probability of .95 in each case. The audiences thus produced are shown below:

Newspaper Audiences in the High
Probability Distribution

	Newspapers		
	Post	Times-Star	Daily Enquirer
Size of High Distribution	178,000	189,000	148,000
(% of Population)	29%	31%	24%
Mean Exposure Probability	.95	.95	.95
Average Audience of the High Distribution (C_1)	169,000	179,500	140,500
(% of Population)	27.5%	29.5%	22.8%

A 1961 national study of newspaper readership[4] found that on an average weekday, 74.2 per cent of the population were primary readers of daily newspapers. Primary readers were defined as those readers who received the newspaper in their homes or bought it directly from a

---

[4]Audits & Surveys Company, Inc., A National Study of Newspaper Reading, Vol. 1, p. 15. The data are for Metropolitan areas of 500,000 or more inhabitants. The data for Sunday readership cited in the next paragraphs are found on page 46 of the same volume.

newsstand; thus, primary readers included all readers except passalong readers. From our estimates above of the proportion of street sales (36 per cent) and home deliveries (64 per cent), we would expect about two-thirds of the primary readers or about 50 per cent (two-thirds of 74.2 per cent) of the population to receive their copies via home delivery. Remarkably, if we assume random duplication between newspapers, these three average audiences imply a net average subscriber audience of 50.5 per cent of the population. This lends some confidence to our estimates.

For the Sunday paper, we have guessed that a much higher percentage of the copies are delivered to households. Whereas we estimated that 64 per cent of the daily papers were delivered to households, we shall set 85 per cent as the proportion of Sunday papers delivered to households. Using 2.23 adults per household, this gives 341,000 adults (180, 150 copies times .85 times 2.23) or 55.2 per cent of the population in the subscriber (high probability) distribution. Since this is such a large proportion of the total population, it seems unlikely that the mean exposure probability will be as large as for the daily papers (.95). Therefore, we have used a mean probability of .85.

The Low Probability Distribution

We have no information about the number of people for each newspaper who are almost never exposed to the particular newspaper. Therefore, we have arbitrarily assumed that one-half of those not in the average audience of a paper are in this low distribution and that its mean probability is 0.05. This implies that these infrequent readers of a given newspaper see an average of one of twenty issues. We have calculated previously the following total average audience values for each newspaper:

Newspaper Average Audience

Post	Times-Star	Enquirer (daily)	Enquirer (Sunday)
293,900	309,300	242,100	369,900
47.6%	50.1%	39.2%	64.3%

Given the size and mean probability for both the high and low distribtuions, the size of the middle distribution is just the remainder of the population and its mean probability is just the remainder of the audience divided by the size. These results are summarized in Table V-3 below.

Table V-3. Distribution Sizes and Average Audiences for Four Newspapers

Newspaper	High Distribution Size	High Distribution Average Probability	High Distribution Average Audience (C_1)	Middle Distribution Size	Middle Distribution Average Probability	Middle Distribution Average Audience (C_1)	Low Distribution Size	Low Distribution Average Probability	Low Distribution Average Audience (C_1)
Post	178,000	.95	169,000	276,600	.419	115,800	161,900	.05	8,100
	29.0%[a]		27.5%[a]	44.8%		18.79%	26.2%		1.31%
Times-Star	189,000	.95	179,500	271,600	.435	118,200	154,400	.05	7,700
	31.0%		29.5%	44.0%		19.15%	25.0%		1.25%
Enquirer									
(Daily)	148,000	.95	140,500	281,600	.326	91,800	187,900	.05	9,400
	24.0%		22.8%	45.6%		14.88%	30.4%		1.52%
Enquirer									
(Sunday	341,000	.85	289,700	167,800	.615	103,100	109,900	.05	5,500
	55.0%		46.7%	27.2%		16.71%	17.8%		.89%

[a]The size and average audience figures are also given as a percentage of the total population.

From the size and mean probability for each distribution, we can calculate the maximum value of the relative accumulation ($C_2^{\,i}/C_1^{\,i}$) for each and the total for the newspaper. This maximum value for the population structured by the three distributions can be compared with the maximum implied simply by the total audience, i.e., the maximum relative accumulation (C_2/C_1) of the same total audience for an unstructured population. These values are shown in Tables V-4 and V-5.

Table V-4. Maximum Two-Period Cumulative Audiences for Each Newspaper Distribution

Distribution	Maximum Two-Period Cumulation as a Percentage of the Total Population for Each Distribution			
	Post	Times-Star	Enquirer (Daily)	Enquirer (Sunday)
High	28.90%	30.90%	23.95%	53.75%
Middle	29.70	29.96	24.90	23.15
Low	2.55	2.44	2.96	1.74
Total	61.15%	63.30%	51.81%	78.64%

Thus we observe that the simple structuring of the population into three distributions with three mean probabilities has significantly decreased the maximum relative accumulation. Comparing the cumulation

Table V-5. Comparison of the Maximum Two-Period Cumulative Newspaper Audiences under the One- and Three-Beta Function Models.

	Maximum Relative Accumulation (C_2/C_1)			
	Post	Times-Star	Enquirer (Daily)	Enquirer (Sunday)
3-Distribution Case	1.284	1.264	1.322	1.224
1-Distribution Case	1.524	1.499	1.608	1.357

percentages above with the audience percentages in Table V-3, we observe that only in the middle distribution is there a significant increase in exposure possible. The high distribution population has been almost completely (85 or 95 per cent) exposed by the first issue, and exposure is so small (5 per cent) for each issue in the low distribution that it fails to accumulate significantly; thus, most of the possible cumulation is restricted to the one-quarter to one-half of the population in the middle distribution. In matters little what figures are chosen for the two-period cumulations in the low and high distributions; the significant figure is that of the middle distribution.

We have only two items of data relating to newspaper cumulation, both from the 1961 newspaper readership study.[5] For all the daily newspapers (with a total

[5]See reference, note [illegible].

average audience of 79.7 per cent of the population), the relative accumulation value is 1.055. This value is quite low, but then the average audience is extremely high. No single newspaper has so high an average audience. We must assume that, at best, this figure is an extreme lower bound for the relative accumulation. For the Sunday newspapers (74.60 per cent average audience) the relative accumulation is 1.045. Since the average audience of the Sunday Enquirer is somewhat lower (64.30 per cent) than the combined newspapers of the study, we expect the Sunday Enquirer's relative accumulation to be somewhat higher. With little guidance from any data, we have arbitrarily set the two-period cumulation for each distribution at the midpoint between the average audience's minimum value and the maximum possible cumulation. Thus, the second issue exposes one-half of those who would be newly exposed if exposure were completely random from issue to issue within the distribution. Table V-6 displays the resulting cumulations.

Distributions and Cumulation for Radio

We consider first the weekday radio broadcasts. The audience data calculated below (Table V-6) and a consideration of the regularity of living patterns during the hours from 7:00 A.M. through 5:00 P.M. seem to invite the common treatment of news broadcast audiences during these

Table V-6. Two-Period Cumulations by Distribution for Each Newspaper

Distribution	Two-Period Cumulation for Each Newspaper			
	Post	Times-Star	Enquirer (Daily)	Enquirer (Sunday)
High	177,200	183,700	144,400	310,200
	28.70%[a]	29.75%	23.38%	50.23%
Middle	149,700	151,700	122,800	123,100
	24.24%	24.56%	19.89%	19.93%
Low	11,900	11,400	13,800	8,200
	1.93%	1.85%	2.24%	1.32%
Total two-period cumulation	338,800	346,800	281,000	441,400
	54.87%	56.16%	45.51%	71.48%
Relative Accumulation	1.162	1.122	1.162	1.113

[a]The percentages are based on the total population.

hours. The percentage of men in the radio audience at these times varies only from 1.48 per cent to 7.19 per cent; thus a large porportion of the men are rarely if at all in these radio audiences. For the women the range is somewhat larger, from 4.62 per cent to 17.77 per cent, and the values somewhat higher than those for the men, as might be expected. The total audiences during this time vary from 3.14 to 12.42 per cent of the population. Therefore, we have decided to treat news broadcasts at these hours as a group.

The Male Audiences

During the hours from 7:00 A.M. through 5:00 P.M., the audience data of Table V-6 shows that not more than 7.19 per cent of the men are ever in the news broadcast audience. At 8:00 A.M. when every station is carrying news and the male listening density is still moderately high, the maximum value audience of the period is reached at 7.19 per cent. We can assume that at any of these hours on any given day the proportion of males not in the radio audience at all is quite large. (Note that the audience figures include a correction for listening outside the home.) Thus we have assigned the men to distributions in the following manner:

1.0 per cent of the men are assigned to a high probability distribution with mean probability of 0.70.

29.0 per cent of the men are assigned to a middle probability distribtuion with a mean probability determined by the average audience of the vehicle.

70.0 per cent of the men are assigned to a low probability distribution with a mean probability of 0.04.

The Female Audiences

For these daytime hours we do have some data about women's listening habits. Lazarsfeld, in a study of women residents of New York, Cleveland, Chicago and Kansas City conducted in the late 1940's, found the following distribution of listenerships among the women:[6]

Inaccessibles	34%
Non-Listeners	24
Story Audience	19
Other Listeners	23
	100%

According to these figures fifty-eight per cent of the women very rarely listen during these hours. The other forty-two per cent do listen. In choosing the distribution sizes for women, we have also noted that the minimum

[6]The study is reported in Paul F. Lazarsfeld and Frank N. Stanton, eds., Communications Research 1948-1949, (New York: Harper & Brothers, 1950), p. 76-85. Although these terms were originally used to describe the morning listening habits of the women, we shall assume that these hold true also for the afternoon. Lazarsfeld indicates in the study (p. 81) that these listening differences are strongly maintained in the afternoon and even to some degree into the evening.

audience for women is as low as 4.62 per cent. The women then are to be distributed as follows:[7]

3.0 per cent of the women are assigned to a high probability distribution with mean probability of 0.70.

39.0 per cent of the women are assigned to a middle probability distribution with a mean probability determined by the average audience of the vehicle.

58.0 per cent of the women are assigned to a low probability distribution with a mean probability of 0.04.

Recalling that the composition of the population is 47.8 per cent male and 52.7 per cent female, we may now produce the three total distributions of listeners and the average audiences accounted for by those of known mean probability:

	Size of Distribution	Mean Probability	Average Audience
High Distribution	12,700 2.05%[a]	0.70	8,900 1.44%[a]
Middle Distribution	211,800 34.27%	To be determined from the individual audience vehicles	
Low Distribution	393,000 63.68%	0.04	15,700 2.55%

[a]The percentages are based on the total population.

[7]Of course, the generation and assignment of probabilities won't strictly follow these male and female distributions; only three distributions are generated by the cumulation routine. We are using this dichotomy (which

The average audience accounted for by the high and low probability distribution is 24,600 or 3.99 per cent of the population. For three of the vehicles (the 9:00 A.M., 10:00 A.M., and 3:00 P.M. news broadcasts) this partial audience is larger than the total vehicle audience. For these three vehicles we have lowered the mean probability of the high distribution to 0.5 and the low distribution to 0.02. The calculated means are presented in Table V-7.

As with the newspaper data above we have arbitrarily set the two-period cumulation for each distribution at the mid-point between its minimum and maximum possible values. Comparing the resulting relative accumulations with those of the four radio programs measured by Politz (Table B-1) we observe that the calculated relative accumulations are slightly lower. However, the evening programs (those measured by Politz) have a potential audience of most of the population; the weekday morning and afternoon programs, however, probably exclude nearly all of the males from the potential audience. Thus the relative accumulation should be slightly lower. The final values are shown in Table V-8.

---

will shortly be united) simply as a heuristic to discover the forms of the three general distributions.

Table V-7. Monday-Friday Morning and Afternoon Audience Distributions and Mean Probabilities

Time	Audience		High Distribution		Middle Distribution		Low Distribution	
	Average Total	Percent of Population	Audience	Probability	Audience	Probability	Audience	Probability
7:00 A.M.	39,200	6.34%	8,890	.70	14,600	.069	15,720	.04
8:00	63,300	10.25	8,890	.70	38,700	.183	15,720	.04
9:00	31,600	5.12	6,350	.50	14,400	.082	7,860	.02
10:00	19,400	3.14	6,350	.50	5,200	.025	7,860	.02
11:00	58,900	9.54	8,890	.70	34,300	.162	15,720	.04
12:00	61,800	10.01	8,890	.70	37,200	.176	15,720	.04
1:00P.M.	76,700	12.42	8,890	.70	52,100	.246	5,720	.04
2:00	---[a]	---	---	---	---	---	---	--
3:00	31,300	5.07	6,350	.50	17,100	.081	7,860	.02
4:00	39,100	6.33	8,890	.70	14,500	.069	15,720	.04
5:00	50,900	8.24	8,890	.70	26,300	.124	15,720	.04

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

Table V-8. Monday-Friday Daytime Radio Two-Period Cumulation by Distribution

Time	Two Period Cumulative Audience of Each Distribution									Total Two-Period Cumulation		Relative Two-period Cumulation
	High Distribution			Middle Distribution			Low Distribution					
	Audience	Percent of Dist.	Percent of Pop.	Audience	Percent of Dist.	Percent of Pop.	Audience	Percent of Dist.	Percent of Pop.	Audience	Percent of Pop.	
7 A.M.	10,250	81.0%	1.66%	21,350	10.1%	3.46%	23,220	5.9%	3.75%	54,800	8.88%	1.402
8	10,250	81.0	1.66	54,400	25.8	8.83	23,220	5.9	3.76	87,950	14.25	1.391
9	7,970	63.0	1.29	25,400	12.0	4.11	11,610	3.0	1.88	44,900	7.28	1.422
10	7,970	63.0	1.29	7,830	3.7	1.27	11,610	3.0	1.88	27,400	4.44	1.415
11	10,250	81.0	1.66	48,630	23.0	7.88	23,220	5.9	3.76	82,060	13.30	1.393
12	10,250	81.0	1.66	52,580	24.9	8.52	23,220	5.9	3.76	86,000	13.94	1.393
1 P.M.	10,250	81.0	1.66	71,700	33.9	11.62	23,220	5.9	3.76	105,200	17.04	1.373
2	---[a]	---	---	---	---	---	---	---	---	---	---	---
3	7,970	63.0	1.29	24,930	11.8	4.04	11,610	3.0	1.88	44,460	7.21	1.422
4	10,250	81.0	1.66	21,350	10.1	3.46	23,220	5.9	3.76	54,800	8.88	1.402
5	10,250	81.0	1.66	37,580	17.8	6.09	23,220	5.9	3.76	71,010	11.51	1.397

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

Evening Listening

For the weekday evening listening distributions, we can find no better data than that of an NORC national survey of fall, 1947.[8] Respondents were asked to estimate the total number of hours that they spent listening in the morning, in the afternoon, and in the evening. Rearranging these figures for evening listening, we find the following distribution of listening:[9]

Percentage of the Population	Average Probability of Listening
13.2%	.000
0.9	.025
3.7	.076
14.6	.150
24.7	.300
20.1	.492
14.6	.700
6.4	.900
1.8	1.000
100.0%	

[8] Paul A. Lazarsfeld and Patricia L. Kendall, Radio Listening in America: The People Look at Radio-Again (New York: Prentice-Hall, Inc., 1948), p. 122.

[9] Assuming that most morning listening occurs in the hours from 7:00 A.M. to noon, the afternoon listening in the hours from noon to 6:00 P.M. and most evening listening in the five hours from 6:00 P.M. to 11:00 P.M., we can convert the respondent's listening estimates into

It is very difficult to use this distribution directly for our purposes since the listening described is an average over all broadcasts at a given time, not simply news broadcasts. (The average audience for weekday evening news broadcasts is never larger than 27 per cent of the population, because at the peak listening hours, one or more of the radio stations is not broadcasting news.)

The data indicate that 18 per cent of the population rarely listen to anything during the evening, about 22 per cent are constantly listening to something, while the remaining 60 per cent listen occasionally to something. Therefore, for the evening news broadcast with the largest audience (27.24 per cent), we have chosen to make the low probability (0.03) distribution 25 per cent of the population. The corresponding high probability (0.75) distribution will equal 20 per cent of the population. For the other vehicles, the lower the average audience, the larger the low distribution and the smaller the high distribution from this base line thus established. Table V-9 shows the resulting distribution sizes and mean probabilities.

---

probabilities of listening at any time. This procedure gives average audiences of 19, 15 and 38 per cent for the morning, afternoon and evening hours, respectively. The Hooper average ratings for Cincinnati during these hours were 18.3, 22.8 and 38.1 per cent respectively. Therefore, we have some confidence in these personal estimates.

Table V-9. Monday-Friday, Evening Radio; Distribution Sizes, Mean Probabilities and Average Audiences.

Time	Average Audience	Distribution Sizes			Distribution Mean Probabilities			Distribution Average Audiences		
		High Dist.	Middle Dist.	Low Dist.	High	Middle	Low	High Dist.	Middle Dist.	Low Dist.
6:00 p.m.	150,300	123,500	339,600	154,400	.75	.156	.03	92,600	53,000	4,630
	24.34%[a]	20%	55%	25%				15.00%	8.58%	.75%
7:00 p.m.	168,200	123,500	339,600	154,400	.75	.209	.03	92,600	71,000	4,630
	27.24%	20%	55%	25%				15.00%	11.50%	.75%
8:00 p.m.	117,200	86,500	314,900	215,100	.75	.146	.03	64,800	46,000	6,480
	18.98%	14%	51%	35%				10.50%	7.45%	1.05%
9:00 p.m.	38,700	12,350	172,900	432,300	.75	.095	.03	9,260	16,400	12,970
	6.27%[b]	2%	28%	70%				1.50%	2.66%	2.10%
10:00 p.m.	---	---	---	---	---	---	---	---	---	---
11:00 p.m.	80,900	61.750	277,900	277,900	.75	.094	.03	46,300	26,100	8,340
	13.10%	10%	45%	45%				7.50%	4.23%	1.35%
12:00 p.m.	21,700	6,175	55,600	555,800	.50	.034	.03	3,090	1,910	16,670
	3.51%	1%	9%	90%				.50%	.31%	2.70%

[a]All percentages are based on the total population.

[b]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

For the two-period cululations of each distribution we have used the procedures as described above for the daytime audiences, with the exception that three-fourths rather than one-half of the maximum additional exposure was allowed. This made the values of the relative two-period cumulation comparable to the Politz radio data cited below (p. 497). The resulting values are tabulated below. (See Table V-10.)

Cumulations for Saturday and Sunday

In estimating the distribution sizes, mean probabilities, and two-period cumulations for these days, we have made the following assumptions:

1. We have assumed that Saturday mornings are like the weekday mornings already calculated, i.e , most of the men and a large proportion of the women are not often available for listening.
2. The remaining Saturday and Sunday vehicles are like weekday evenings, i.e., the men are available and the programs have relatively high relative two-period cumulations.

We note that for each week in the time period of the simulation, each weekday vehicle will run five times compared to one run of each weekend vehicle. Thus, the weekday vehicles should have a much more important effect on exposure. The weekend vehicle data are shown in Tables V-11 to V-14 below.

Table V-10. Monday-Friday Evening Radio Two-Period Cumulative Audience by Distribution

Time	Two-Period Cumulative Audience of Each Distribution									Total Two-Period Cumulation		Relative Two-Period Cumulation
	High Distribution			Middle Distribution			Low Distribution					
	Audience	Percent of Dist.	Percent of Pop.	Audience	Percent of Dist.	Percent of Pop.	Audience	Percent of Dist.	Percent of Pop.	Audience	Percent of Pop.	
6 p.m.	109,900	89.0%	17.80%	86,600	25.5%	14.03%	8,200	5.3%	1.33%	204,800	33.16%	1.362
7	109,900	89.0	17,80	113,100	33.3	18.32	8,200	5.3	1.33	231,300	37,45	1.375
8	76,900	89.0	12.46	75,600	24.0	12.24	11,500	5.3	1.89	164,200	26.59	1.401
9	11,000	89.0	1.78	25,500	15.9	4.45	22,900	5.3	3.71	61.400	9.94	1.585
10	---[a]	---	---	---	---	---	---	---	---	---	---	---
11	55,000	89.0	8.90	43,900	15.8	7.11	14,700	5.3	2.39	113,600	18.40	1.405
12	4,200	69.9	.69	3,300	5.9	.53	29,500	5.3	4.77	37,000	5.99	1.707

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

Table V-11. Distribution Sizes for Saturday Morning News Broadcasts

Distribution	Distribution Size	Per Cent of Population
High	12,700	2.05%
Middle	211,800	34.27
Low	393,000	63.68

Table V-12. Saturday Morning News Broadcast Distribution Audiences and Mean Probabilities

Time	Average Audience		High Distribution		Middle Distribution		Low Distribution	
	Total	Per Cent of Population	Audience	Mean Probability	Audience	Mean Probability	Audience	Mean Probability
7:00 A.M.	39,800	6.45%	8,890	.70	15,200	.072	15,720	.04
8:00	60,900	9.86	8,890	.70	36,300	.172	15,720	.04
9:00	49,900	8.08	8,890	.70	25.300	.120	15,720	.04
10:00	50,000	8.10	8,890	.70	25,400	.120	15,720	.04
11:00	60,900	9.86	8,890	.70	36,300	.172	15,720	.04
12:00	77,900	12.62	8,890	.70	53,300	.252	15,720	.04

Table V-13. Saturday Afternoon and Evening News Broadcasts: Distribution Sizes, Mean Probabilities, and Average Audiences

Time	Average Audience		Distribution Sizes					
	Total	% of Pop.	High Dist.	% of Pop.	Middle Dist.	% of Pop.	Low Dist.	% of Pop.
1:00P.M.	67,800	10.98%	30,900	5.0%	216,100	35.0%	370,500	60.0%
2:00	19,900	3.22	3,100	0.5	58,700	9.5	555,800	90.0
3:00	70,900	11.48	30,900	5.0	216,100	35.0	370,500	60.0
4:00	92,500	14.98	61,750	10.0	247,000	40.0	308,700	50.0
5:00	101,200	16,38	86,500	14.0	314,900	51.0	216,100	35.0
6:00	171,700	27.78	123,500	20.0	339,600	55.0	154,400	25.0
7:00	176,700	28.61	123,500	20.0	339,600	55.0	154,400	25.0
8:00	81,800	13.25	61,750	10.0	247,000	40.0	308,700	50.0
9:00	---[a]	---	---	---	---	---	---	---
10:00	15,700	2.54	3,100	0.5	58,700	9.5	555,800	90.0
11:00	100,300	16.25	86,500	14.0	314,900	51.0	216,100	35.0
12:00	19,100	3.09	3,100	0.5	58,700	9.5	555,800	90.0

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were schedules, or both.

Table V-13, (continued).

Time	Distribution Mean Probabilities			Distribution Average Audiences					
	High Dist.	Middle Dist.	Low Dist.	High Dist.	% of Pop.	Middle Dist.	% of Pop.	Low Dist.	% of Pop.
1:00 P.M.	.75	.155	.03	23,200	3.75%	33,500	5.43%	11,120	1.80%
2:00	.50	.123	.02	1,540	.25	7,200	1.17	11,120	1.80
3:00	.75	.169	.03	23,200	3.75	36,600	5.93	11,120	1.80
4:00	.75	.150	.03	46,300	7.50	36,900	5.98	9,260	1.50
5:00	.75	.095	.03	64,800	10.50	29,800	4.83	6,590	1.05
6:00	.75	.219	.03	92,600	15.00	74,300	12.03	4,630	.75
7:00	.75	.234	.03	92,600	15.00	79,400	12.86	4,630	.75
8:00	.75	.106	.03	46,300	7.50	26,200	4.25	9,260	1.50
9:00	---[a]	---	---	---	---	---	---	---	---
10:00	.50	.052	.02	1,540	.25	3,000	..49	11,120	1.80
11:00	.75	.092	.03	64,800	10.50	29,000	4.70	6,480	1.05
12:00	.50	.109	.02	1,540	.25	6,500	1.04	11,120	1.80

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

Table V-14. Saturday News Broadcast Two-Period Cumulative Audiences

Time	Two-Period Cumulative Audience of Each Distribution									Total Two-Period Cumulation		Relative Two-Period Cumulation
	High Distribution			Middle Distribution			Low Distribution					
	Cumulation	% of Dist.	% of Pop.	Cumulation	% of Dist.	% of Pop.	Cumulation	% of Dist.	% of Pop.	Total	% of Pop.	
7:00A.M.	10,250	81.0%	1.66%	22,400	10.6%	3.63%	23,220	5.9%	3.76%	55,900	9.05%	1.403
8:00	10,250	81.0	1.66	51,400	24.3	8.33	23,220	5.9	3.76	84,900	13.75	1.395
9:00	10,250	81.0	1.66	36,600	17.3	5.93	23,220	5.9	3.76	70,100	11.35	1.405
10:00	10,250	81.0	1.66	36,600	17.3	5.93	23,220	5.9	3.76	70,100	11.35	1.401
11:00	10,250	81.0	1.66	51,400	24.3	8.33	23,220	5.9	3.76	84,900	13.75	1.395
12:00	10,250	81.0	1.66	72,200	34.6	11.86	23,220	5.9	3.76	106,700	17.28	1.369
1:00P.M.	27,500	89.0	4.45	54,700	25.3	8.86	19,640	5.3	3.18	101,800	16.49	1.502
2:00	2,160	69.0	.35	12,000	20.4	1.94	19,450	3.5	3.15	33,600	5.44	1.688
3:00	27,500	89.0	4.45	59,200	27.4	9.59	19,640	5.3	3.18	106,300	17.22	1.500
4:00	55,000	89.0	8.90	60,800	24.6	9.84	16,360	5.3	2.65	132,100	21.39	1.428
5:00	76,900	89.0	12.46	50,400	16.0	8.16	11,490	5.3	1.86	138,800	22.48	1.372
6:00	109,900	89.0	17.80	117,900	34.7	19.09	8,210	5.3	1.33	236,000	38.22	1.375
7:00	109,900	89.0	17.80	125,000	36.8	20.24	8,210	5.3	1.33	243,100	39.37	1.376
8:00	55,000	89.0	8.90	43,700	17.7	7.08	16,360	5.3	2.65	115,000	16.63	1.406
9:00	---[a]	---	---	---	---	---	---	---	---	---	---	---
10:00	2,160	69.0	.35	5,200	8.9	.85	19,450	3.5	3.15	26,900	4.35	1.711
11:00	76,900	89.0	12.46	48,800	15.5	7.91	11,490	5.3	1.86	137,300	22.23	1.369
12:00	2,160	69.0	.35	10,700	18.2	1.73	19,450	3.5	3.15	32,300	5.23	1.691

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

Table V-15. Sunday News Broadcasts: Distribution Sizes, Mean Probabilities, and Average Audiences

Time	Total Audience		Distribution Sizes					
	Total	% of Pop.	High Dist.	% of Pop.	Middle Dist.	% of Pop.	Low Dist.	% of Pop.
8:00 A.M.	31,800	5.15%	3,100	0.5%	58,700	9.5%	555,800	90.0%
9:00	24,200	3.92	3,100	0.5	58,700	9.5	555,800	90.0
10:00	21,200	3.43	3,100	0.5	58,700	9.5	555,800	90.0
11:00	54,100	8.76	18,400	3.0	183,800	30.0	410,400	67.0
12:00	122,900	19.90	86,500	14.0	314,900	51.0	216,100	35.0
1:00 P.M.	97,500	15.79	61,750	10.0	247,000	40.0	308.700	50.0
2:00	15,600	2.53	3,100	0.5	58,700	9.5	555,800	90.0
3:00	53,200	8.62	18,400	3.0	183,800	30.0	410,400	67.0
4:00	19,200	3.11	3,100	0.5	58,700	9.5	555,800	90.0
5:00	25,400	4.11	3,100	0.5	58,700	9.5	555,800	90.0
6:00	99,300	16.08	86,500	14.0	314,900	51.0	216,100	35.0
7:00	14,500	2.35	3,100	0.5	58,700	9.5	555,800	90.0
8:00	25,300	4.10	3,100	0.5	58,700	9.5	555,800	90.0
9:00	15,800	2.56	3,100	0.5	58,700	9.5	555,800	90.0
10:00	---[a]	---	---	---	---	---	---	---
11:00	107,900	17.47	86,500	14.0	314,900	51.0	216,100	35.0
12:00	25,700	4.16	3,100	0.5	58,700	9.5	555,800	90.0

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

Table V-15 (continued).

Time	Dist. Mean Probabilities			Distribution Average Audiences					
	High Dist.	Middle Dist.	Low Dist.	High Dist.	% of Pop.	Middle Dist.	% of Pop.	Low Dist.	% of Pop.
8:00 A.M.	.50	.326	.02	1,540	.25%	19,100	3.10	11,120	1.80%
9:00	.50	.197	.02	1,540	.25	11,500	1.87	11,120	1.80
10:00	.50	.145	.02	1.540	.25	8,500	1.38	11,120	1.80
11:00	.75	.150	.03	13,800	2.25	27,800	4.50	12,300	2.01
12:00	.75	.164	.03	64,800	10.50	51,600	8.35	6,480	1.05
1:00 P.M.	.75	.170	.03	46,300	7.50	41,900	6.79	9,260	1.50
2:00	.50	.051	.02	1,540	.25	3,000	.48	11,120	1.80
3:00	.75	.145	.03	13,800	2.25	26,900	4.36	12,300	2.01
4:00	.50	.112	.02	1,540	.25	6,500	1.06	11,120	1.80
5:00	.50	.217	.02	1,540	.25	12,700	2.06	11,120	1.80
6:00	.75	.089	.03	64,800	10.50	28,000	4.53	6,480	1.05
7:00	.50	.032	.02	1,540	.25	1,800	.30	11,120	1.80
8:00	.50	.216	.02	1,540	.25	12,700	2.05	11,120	1.80
9:00	.50	.054	.02	1,540	.25	3,100	.51	11,120	1.80
10:00	---[a]	---	---	---	---	---	---	---	---
11:00	.75	.116	.03	64,800	10.50	36,600	5.92	6,480	1.05
12:00	.50	.222	.02	1,540	.25	13,000	2.11	11,120	1.80

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

Table V-16. Sunday News Broadcast Two-Period Cumulative Audiences

Time	Two-Period Cumulative Audience of Each Distribution									Total Two-Period Cumulation		Relative Two-Period Cumulation
	High Distribution			Middle Distribution			Low Distribution					
	Cumulation	% of Dist.	% of Pop.	Cumulation	% of Dist.	% of Pop.	Cumulation	% of Dist.	% of Pop.	Total	% of Pop.	
8:00 A.M.	2,160	69.0%	.35%	28,800	49.1%	4.66%	19,450	3.5%	3.15%	50,400	8.16%	1.584
9:00	2,160	69.0	.35	18,500	31.6	3.00	19,450	3.5	3.15	40,100	6.50	1.658
10:00	2,160	69.0	.35	14,000	23.8	2.26	19,450	3.5	3.15	35,600	5.76	1.679
11:00	16,500	89.0	2.67	45,600	24.6	7.38	21,900	5.3	3.55	84,000	13.60	1.553
12:00	76,900	89.0	12.46	84,100	26.7	13.62	11,490	5.3	1.86	172,500	27.94	1.404
1:00 P.M.	55,000	89.0	8.90	68,200	27.6	11.04	16,360	5.3	2.65	139,500	22.59	1.431
2:00	2,160	69.0	.35	5,100	8.7	.83	19,450	3.5	3.15	26,700	4.33	1.711
3:00	16,900	89.0	2.67	44,100	23.8	7.14	21,900	5.3	3.55	82,500	13.36	1.550
4:00	2,160	69.0	.35	11,000	18.6	1.77	19,450	3.5	3.15	32,500	5.27	1.695
5:00	2,160	69.0	.35	20,300	34.5	3.28	19,450	3.5	3.15	41,900	6.78	1.650
6:00	76,900	89.0	12.46	47,200	15.0	7.65	11,490	5.3	1.86	135,700	21.97	1.366
7:00	2,160	69.0	.35	3,200	5.5	.52	19,450	3.5	3.15	24,800	4.02	1.711
8:00	2,160	69.0	.35	20,000	34.1	3.24	19,450	3.5	3.15	41,600	6.74	1.644
9:00	2,160	69.0	.35	5,400	9.2	.87	19,450	3.5	3.15	27,000	4.37	1.707
10:00	---[a]	---	---	---	---	---	---	---	---	---	---	---
11:00	76,900	89.0	12.46	60,800	19.3	9.84	11,490	5.3	1.86	149,200	24.16	1.383
12:00	2,160	69.0	.35	20,600	35.2	3.34	19,450	3.5	3.15	42,200	6.84	1.558

[a]The news broadcast audience at this hour is insignificantly small, either because there were few listeners at all at this hour, or because few news broadcasts were scheduled, or both.

CHAPTER VI

CONSTRUCTING THE MEDIA SYSTEM IV: THE AUDIENCE DUPLICATIONS AND ASSIGNMENT OF EXPOSURE PROBABILITIES

Allocating Probabilities to Population Subgroups

In Chapter I we introduced the basic ideas relating to the allocation of the probabilities generated from the cumulation statistics to the cells defining the population subgroups. Before we present the treatment of duplication in the model, we need to describe in detail this allocation procedure.

In the first link of the simulation, we generated the simulation population and the audiences for each communication vehicle distributed over the population types defined by the population dimensions. The ratio of the audience in a cell to the population of that cell gave a resulting mean probability of exposure for the particular cell. In the second link of the simulation, we introduced, for each vehicle, the cumulation information (described in Chapter V), for each of three distributions. Then, for each distribution, we calculated the parameters of the beta function and created a number of probabilities equal

to the distribution size and distributed at intervals from 0.0 to 1.0 in approximately the density of the beta function. (Of course, the discreet probabilities do not exactly duplicate the beta function; therefore, if the average audience is slightly different from that required, the simulation makes a slight correction in these discreet probabilities.) Finally, for each vehicle we have three distributions of discreet probabilities, the number of probabilities in each distribution equal to the number of simulated persons in that distribution, and the probabilities chosen in such a way as to reproduce the cumulation data for each distribution and for the vehicle as a whole. The next step is to allocate these probabilities to the cells of the simulation population in such a way as to reproduce each cell's average audience for the vehicle.

How shall we decide which probabilities are allocated to which cells? First, the number allocated to any cell must equal the population size for the cell. Second, the average of the allocated probabilities should equal the vehicle average audience in the cell. If we have only <u>two</u> distributions of probabilities, if we draw randomly without replacement from the distributions, and if we are willing to modify the second condition such that the <u>expected value</u> (over many draws) of the average of the allocated probabilities be equal to the cell mean, then we can specify for each cell the exact number of probabilities

to be drawn randomly from the high and low distributions. However, if we have three distributions, we need another condition to specify exactly the number of probabilities for each cell to be drawn from each distribution. We have actually specified a function which relates the cell mean, the average of the probabilities of each of the three distributions, and the number of probabilities to be drawn from each distribution, in order to provide the third condition to solve the equations.[1] However, this method of allocating probabilities, while sufficiently accurate for cells with a large number of probabilities, (in that it will closely reproduce the cell mean), is not accurate enough for small cells. Therefore, we have developed another method for allocating probabilities which takes account of the cell mean to be reproduced after each draw of a probability. This method is used in the simulation as presently programmed; it is described in the succeeding paragraphs.

Let us call the three distributions of probabilities produced for each vehicle from the cumulation data the low, the middle, and the high probability distributions. Of

---

[1]This solution to the problem is discussed in detail in John F. Kramer, The Three Distributions Case: How to Choose the Proper Number of Probabilities for Each Distribution When there are Three Distributions, COMCOM/Simulation Memo #33, January 27, 1966. It turns out that any function lying within a specified convex set and passing through two of the corners of that set will provide a satisfactory solution.

course, the probabilities in each of these distributions may range from 0.0 to 1.0. However, the average value of the probabilities in the low distribution will be the lowest of all the distributions with correspondingly higher average probabilities for the other two distributions. We define boundaries between the distributions at the following points: the boundary between the low and middle distributions is the average of the mean probability of the low distribution and the mean probability of the middle distribution; the boundary between the middle and high distributions is the average of the mean probability of the middle distribution and the mean probability of the high distribution. Now we look at the first cell in the simulation. The cell mean for this first cell falls in the range, as defined by these boundaries, either of the low, the middle, or the high distribution. For instance, if the boundary between the low and the middle distribution is .30 and the mean of the cell is .24, then the cell mean at this point falls in the range of the low distribution. Therefore, in order to best approximate this cell mean, we draw randomly from the probabilities of the low distribution and assign the resulting probability to this cell. At the same time, we eliminate this probability from the list of low probabilities. This low probability may, in fact, take on any value between 0.0

and 1.0; however, in general it will have a value near the average value for the low distribution.

At this point there remains one less probability to be assigned to this cell. The required average of these remaining probabilities is not, however, equal to the cell mean. If the first probability assigned happened, by chance, to be lower than the cell mean, then the average of the remaining probabilities must be higher than the cell mean in order that the average probability for the cell be equal to the cell mean. On the other hand, if the assigned probability happened to be larger than the cell mean, the average of the remaining probabilities must be smaller than the cell mean. Therefore, we calculate this new required average probability for the yet-to-be assigned probabilities and again draw randomly from the appropriate distribution implied by this new average, deleting the new probability from the list after it has been assigned to the cell. This process is repeated probability-by-probability and cell-by-cell. If a list is exhausted, the probabilities are drawn from the distribution which is closest to the mean of the probabilities to be drawn. Eventually, all the probabilities are assigned in this way to all of the cells. This method of random assignment does not exactly reproduce the cell means as required from the first link of the simulation; however, we recall that the cell means produced by the parameter

estimation process are themselves only estimates of the true values. Moreover, we are not likely to analyze the simulation output for a single cell (unless it is a very large cell), but for groups of cells, and the larger the number of probabilities in the cell or groups of cells, the more likely is the assignment process to reproduce closely the mean probability required from the first link of the simulation.[2]

Thus, upon completion of the second link of the simulation, we have, for each vehicle, probabilities of exposure assigned to each cell of the population in such a way as to closely reproduce the cell mean audience for the vehicle. These probabilities, however, have not yet been assigned to specific individuals within the population cells. In making this assignment we must consider the audience duplication between vehicles.

---

[2]While writing this description, an improvement to this assignment process becomes evident. In addition to recalculating after each draw the average required for the remaining probabilities one could also recalculate the average for each distribution after the probability drawn has been deleted. Then the new distribution means and cell mean would govern the choice of the distribution from which the next probability is to be drawn. Also, the boundaries between the distributions could be defined by some sort of weighted average of the mean distribution probabilities. Finally, we note that the cells are processed in their logical order. Perhaps it would be better to process them according to size, from largest to smallest, since the smaller cells are not usually as important and since the few draws required to fill them will not afford much opportunity for recalculation of means and correction anyway.

Duplication in the Vehicle Audiences

As we have mentioned before, there are two important ways in which we specify structure in the mass media simulation. The first of these is in the organization of the probabilities of exposure for a particular vehicle. The differences in these probabilities result in some people being almost constantly exposed to the vehicle, while other people are almost never exposed to the vehicle. These two extremes combine with people who are sometimes exposed to the vehicle to produce a net audience which is just equal to the mean audience figures which we have arrived at from our survey data and other considerations.

The second way in which structure is programmed into the simulation is through the duplication among the vehicle audiences. Time and time again, audience studies have shown that those people who have a high probability of exposure to one vehicle generally have a high probability of exposure to other vehicles; that in fact there exists a syndrome which we might label the high media-consumption syndrome. Conversely, there also exists a low media-consumption syndrome; that is, there are individuals who have very low probabilities of exposure to any and all of the vehicles in the mass media. Therefore, just as there are individuals who almost always see every issue or broadcast of a given vehicle, there are also individuals--

and to a large extent, the same individuals--who are highly exposed to a wide range of vehicles. Even among those individuals who are exposed to a wide range of vehicles, however, we often find relatively low duplication between two vehicles which serve more or less equivalent functions, e.g., Newsweek and Time magazines. Since the contents of these magazines tend to be quite similar, a reader of one magazine is not likely to be a reader of the other, even though the demographic characteristic of the two audiences may be quite similar.

The simulation accounts for this non-random duplication between vehicle audiences in two ways. The first involves the cell mean probabilities of exposure for each population type for each vehicle. The cell means cause non-random duplication in the following manner: assume that certain people are more likely to be exposed to evening news broadcasts than are other people and that these are people of higher education. If these people are also more likely to be exposed to front-page newspaper news, then the cell means for people of higher education will be higher, both for evening news broadcasts and for newspaper readership. The product of these two probabilities, which in general will be large because each of the two probabilities is itself high, gives the expected audience duplication if the assignment of probabilities for the two vehicles reproduces the cell means and

otherwise is done randomly across members of that particular cell. Thus, the demographic distribution of the audiences of the several vehicles does itself account for much of the non-random duplication of audiences among those vehicles. In fact, in most audience studies, the factors which are used to explain the phenomenon of wide media exposure among a given group of individuals are generally the demographic and sociological characteristics of that group. For the case of non-random duplication which is <u>not</u> accounted for by the cell means, we must make a non-random assignment of probabilities to individuals <u>within</u> cells.

Within-Cell Non-Random Probability Assignments

The third and fourth links of the simulation perform the calculations which enable the simulation to take some account of duplication which is not accounted for by the demographic characteristics as implied in the cell means. For those vehicles for which the duplicated audience is known or assumed, that audience is entered into the simulation, and a matrix of the duplicated audiences between every possible pair of vehicles is produced. From the cell means produced by the first link of the simulation, we also generate a corresponding matrix of <u>expected</u> duplications, giving <u>random</u> within-cell assignments of probabilities. Corresponding elements of the two matrices

of random and real-world duplications among vehicles are then compared, and if a difference of more than five percent of the real-world figure is found, the square of this difference is entered in a third matrix of squared differences of duplications. For those vehicles about which there is no *a priori* assumption about duplication or real-world duplication information, the value entered into the third matrix is 0.0. We have produced, then, a matrix of the squared differences between the known or assumed empirical duplication and the random (at the cell level) duplication implied by the simulation.

Any correction we make now for the duplication should, in theory, be made simultaneously for all vehicles having significant non-random duplication, i.e., it is not sufficient to correct the duplication between vehicle A and vehicle B, and then between vehicle A and vehicle C, since duplications between A and B, A and C, *and* B and C, *all* may require correction. We have not been able to discover an algorithm which would allow the correction of duplication for all the possible cases simultaneously. We use instead a correction which will consider up to five vehicles at a time. The algorithm, suggested by Robert P. Abelson,[3] groups the vehicles into overlapping sets of five such that each set is closely bound together

---

[3]Robert P. Abelson, Department of Psychology, Yale University, July 25, 1964, personal letter.

in the sense that the duplications among the five vehicles are quite non-random. It chooses as the first four vehicles in the first set, those four for which the sums of the squared differences between the real world and expected (based on audience-derived cell means) duplication over all vehicles in the matrix, are largest. The fifth vehicle in the set is that vehicle for which the sum over the first four vehicles is largest, i.e., the vehicle with the largest interaction with the first four. To form the next group of five, the algorithm finds which of the remaining vehicles has the largest interaction with the first group. It then removes from the group the vehicle with the smallest interaction with the newly identified vehicle and adds the latter, forming a new group of five vehicles. Thus, we have a second set of five vehicles, four of which were included in the first set of five vehicles. We proceed in this way until all the vehicles for which there is duplication information are included in a set of linked vehicles.[4]

The fourth link of the simulation continues calculation for the assignment of probabilities in order to reproduce non-random duplication. At first the vehicles are treated in pairs, each pair consisting of two of the

---

[4]This algorithm is discussed in detail by Herbert J. Selesnick in Link: A Program that Groups Media Among Which There Exists a Non-Random Duplication, COMCOM/Simulation Memo #17, August 26, 1964.

vehicles included in one of the sets of five linked vehicles. Again we treat the problem in the expected value sense and ask: How shall we make the assignment of probabilities for the two vehicles so as to reproduce the known duplication? For each member of the population, the pair of probabilities for the two vehicles can be categorized by its distribution (high, middle, low) on vehicle A and vehicle B: therefore, there are nine possible categories of pairs ranging from low distribution probabilities for both vehicle A and vehicle B. Figure VI-1 below indicates the nine possibilities.

		Distribution on Vehicle A		
		High	Middle	Low
Distribution on Vehicle B	High	x_{11}	x_{12}	x_{13}
	Middle	x_{21}	x_{22}	x_{23}
	Low	x_{31}	x_{32}	x_{33}

Figure VI-1. Logical Distribution Types for Pairs of Probabilities.

If we let x_{ij} represent the number of people (or assignment pairs) in the ij^{th} cell, let PL_A, PM_A, and PH_A represent the average probability of the low, middle and high distributions of vehicle A, and PL_B, PM_B, PH_B represent the corresponding averages for vehicle B, then the

expected audience duplication is given by the equation:

$$\begin{aligned} \text{Expected Duplication} = {} & x_{11}\cdot PH_A\cdot PH_B + x_{12}\cdot PM_A\cdot PH_B \\ & + x_{13}\cdot PL_A\cdot PH_B + x_{21}\cdot PH_A\cdot PM_B \\ & + x_{22}\cdot PM_A\cdot PM_B + x_{23}\cdot PL_A\cdot PM_B \\ & + x_{31}\cdot PH_A\cdot PL_B + x_{32}\cdot PM_A\cdot PL_B \\ & + x_{33}\cdot PL_A\cdot PL_B \end{aligned}$$

In addition, the total number in each distribution for each vehicle is known, giving five independent equations of the form

$$\text{Number in High Distribution for Vehicle A} = x_{11} + x_{21} + x_{31}.$$

Any solution of these six equations involving the number of probability pairs will produce the desired cumulation between two vehicles. It is possible that there is no solution to these equations, i.e., that the audience duplication required empirically is either too large or too small to be reproduced in this expected value sense, given the distribution means and sizes for the two vehicles. However, this would be unlikely if the means of the low and high distributions were quite different for both of the vehicles. In any event, a calculation is made to determine the maximum and minimum values of the

duplication which can be reproduced by the distributions. The calculation is made according to the following theorem:

> Theorem: Given that we have two sets of N positive numbers. Consider the sum of the N products formed by multiplying a number from one set by a number from the other set. (Each number is included in only one product.) Then the largest value of the sum is achieved in the following way: multiply the highest number in the first distribution by the highest number in the second distribution, the second highest number in the first distribution by the second highest number in the second distribution, etc. The lowest value for the summed products is formed by multiplying the highest number in the first distribution by the lowest number in the second distribution, the second highest number in the first distribution by the next-to-lowest number in the second distribution, etc.

Using this theorem, the calculation of the largest and smallest values of the duplication which can possibly be reproduced is straightforward, given the means and sizes of the three distributions of each of the two vehicles. If we find that the actual duplication falls outside this range, the values for the number of people in the ninefold categories are set at those values which come closest to reproducing the empirical duplication, e.g., if the empirical duplication is too high to be reproduced, the values of the x_{ij} are set (according to the theorem above) to produce the maximum possible duplication. If the empirical duplication falls within the range of possible duplications, then a linear programming algorithm is simply used to find a feasible solution to

the set of five equations and nine unknowns. Thus, for each pair of linked vehicles, a table of the nine x_{ij} is produced.

Just as there are nine probability-pair types for each pair of vehicles, there are for each group of five vehicles (each vehicle having three distributions) 3^5 or 243 possible probability-quintuple types (x_{ijklm}) based on the probability distribution for each vehicle. If we can calculate values for the 243 X_{ijklm} which are consistent with the nine x_{ij} for each of the ten pairs of vehicles in a set of five vehicles (given that the duplication information was there for all pairs of the vehicles), then an assignment of probabilities according to the 243 x_{ijklm} would, on the average, exactly account for all the pairwise duplication in the set of five vehicles. By now it should be clear how we go from a set of subtables to the overall grand table; we use the smoothing iteration to estimate the frequencies in the five-dimensional table from the ten possible two-dimensional tables. It sometimes happens, in performing this iteration, that the ten two-dimensional tables have inconsistencies. These inconsistencies do not necessarily arise because the data is inconsistent, but because the particular feasible solutions arrived at by the linear programming algorithm are inconsistent. At any rate, it is sometimes necessary to omit one or more of the two-dimensional tables in

order that the iteration converges. Then the output from this fourth link of the simulation is the proportion of the probability-quintuple assignments (or of the population) which falls into each of the 243 types, for each set of five vehicles. This information, in the form of percentages, is then passed on to the fifth link of the simulation which makes the actual assignments of probabilities to individuals.

The fifth link of the simulation does the actual assignment of probabilities to individuals for all the vehicles in the media system. The probabilities for those vehicles which are not involved in any way with other vehicles, i.e., those vehicles which do not appear in any of the linked groups either because they have random duplication with other vehicles or because there are no data about the duplication, are assigned randomly to individuals within each cell. Also, the first probabilities for the first vehicle of the first group of five linked vehicles are assigned randomly. For the second vehicle, we proceed in the following fashion: the 243 percentages are summed over the three vehicles in which we are not yet concerned to get the nine percentages, (just the nine x_{ij} for the first two vehicles of the group, expressed as percentages), which describe the required probability-pair types for the first and second vehicles. These x_{ij} express the relationship between the two vehicles

required in the entire population to reproduce the empirical duplication. In attempting to reproduce this duplication for the cell, however, we are constrained by the limited number of people in any cell. Often, we may have even fewer people in the cell than we have probability assignment types, especially when the number of types is 243. In this case, we obviously cannot reproduce the required assignments; we can only approximate it.

Also, the probabilities for each of the vehicles have already been allocated to the cell. For any vehicle, the probability distribution at the cell level will not, in general, be identical with the distribution for the entire population which was, of course, used in generating the 243 x_{ijklm}. We attempt, at the cell level, to approximate the percentages of the 243 types by performing the smoothing iteration again. The subtables used in this process for the case of the first and second vehicles in the first set are simply the numbers of probabilities in the cell from each distribution for each of the two vehicles. Thus, we have two three-value tables for each vehicle which must be put together as closely as possible to resemble the ninefold table of combinations implicit in the 243 percentages. The matrix of nine combinations is <u>initialized</u> with the exact values implicit in the 243 percentages and then the iteration is performed to make the matrix conform to the two three-value marginals. The

iteration usually produces non-integral values for the nine types which must be made integral (since one can assign only integral numbers of probabilities) and yet conform to the marginals.

First, we produce integral values by a Monte Carlo rounding of each of the fractions. Then, we check each of the rows and columns against the marginal sums and, if errors are found, we add or subtract from the column where the error occurs and correct the corresponding row value. In this way, we reproduce as closely as possible (given the constraints of the probabilities already drawn for the individual cell) the assignments required over the entire population in order to reproduce the empirical duplication. This process is extended in an obvious way for duplication between the first and second vehicle as a pair and a third vehicle in the group of five, etc.

The assignment process must at many points be sacrificed to the more important preservation of cell means by maintaining the probabilities for each vehicle within the cells. We also have not attempted, at any point, to look at <u>individual</u> probabilities, but rather have treated distributions of probabilities in terms of their known averages. Thus, the assignment process cannot be expected to fully achieve the required empirical duplication; however, it does make an attempt to bias the assignment of

probabilities in the direction required to more nearly approximate the pairwise real-world vehicle duplication.

Estimating the Empirical Duplications

We do not know, in general, how much of the duplication is accounted for by the demographic factors and how much duplication remains due to other factors which have not been specified as dimensions of the simulation population. However, the simulation itself may help to get a firmer grasp of the magnitude of the effects of the non-cell-defining, but audience influencing, characteristics, since given a known duplication it will examine the duplication which would be produced by random within-cell assignment, and measuring the difference between the random assignment-produced duplication and the empirical duplication, it will decide whether or not a non-random assignment is required. Therefore, the simulation actually makes a decision, where the empirical duplication dataare available, about the strength of the effects of the non-cell-defining variables. In the next section, we estimate duplication of audiences for those vehicles for which we can make reasonable estimates of that duplication. For example, the duplication among the three daily newspapers was measured to some degree by the NORC survey which was used to generate the original audience estimates for those vehicles. On the other hand, for the

radio audiences we have no direct, empirical measurement of the audience duplication. There do exist some Hooper and Nielsen data which give indications of likely audience duplication for consecutive programs or, for example, for programs in the same time spot on weekday evenings. In these cases, we will attempt estimates of the audience duplication. It remains to be seen whether these duplication estimates differ substantially enough from the duplication implied by the cell means to require the simulation to actually make non-random probability assignments within cells. As for the duplication of audiences between the newspapers and the various radio broadcasts, there seems to be very little data other than the general knowledge that people who are highly exposed to one vehicle and in particular specific kinds of news in one vehicle, are quite likely to be exposed to other vehicles, and that same kind of news in other vehicles. However, as was stated above, there is little reason to believe that these duplication effects are not rather well explained by the population dimensions, i.e., the demographic and sociological characteristics of the audience, rather than by factors which have not been considered. Therefore, in these cases, we will not attempt to derive an audience duplication. Let us proceed now to the audience duplication among the three weekday newspapers.

Audience Duplication Among the Three Daily Newspapers

It will be recalled from Chapter III (p.) that the answers to the NORC question concerning readership of the daily newspapers are to be interpreted as follows: a "no" answer for any paper is considered to be exactly that, i.e., no readership. However, a "yes" answer is taken to mean a probability of 0.8855 of exposure to that newspaper on that given day. From this and the duplication and triplication as measured by the NORC survey, we are able to arrive at the duplication between any two of the daily newspapers. For example, of those people who claim exposure to both the Post and Times-Star, we consider that a proportion equal to the square of the probability of exposure were actually exposed to both newspapers on any two random occasions. There is also, however, another group of people who may have been exposed to both newspapers, namely those people who claim exposure to all three newspapers but who, according to our probability measure, were exposed to only two of the three. This is the number claiming exposure to three newspapers multiplied by the square of the probability and by the factor 1 minus the probability. By this method, we have estimated the probability of exposure to any two newspapers (see Table VI-1).

We note that the corrected duplications are nearly

Table VI-1. NORC and Corrected Empirical Audience Duplication Between Pairs of Daily Newspapers as Percentages of the Total Population.

Duplication	Newspaper Pair		
	Enquirer-Post	Enquirer-Times-Star	Post-Times-Star
Raw NORC	9.62%	11.91%	5.04%
Corrected by the Probability Factor .8855	9.52	10.60	5.22

equal to the raw duplications. Although the raw duplications were somewhat diminished (by a factor of about 78 per cent) by our correction, it was also increased by the approximately 9 percent of those who claimed reading of all three newspapers, but who were assigned to the category of readers of only two of the newspapers. Thus, we see that in fact the duplication for the Post and Times-Star has been slightly increased from the raw duplication. In passing, we note also that the smallest duplication is between the Post and Times-Star. This is to be expected since the Post and Times-Star are both evening newspapers and the Enquirer is a morning newspaper.

For the case of the Sunday Enquirer, we shall make no estimate of the duplication between it and any other vehicles since there are no data available. It would seem reasonable to estimate that the duplication between

the Sunday *Enquirer* and the daily *Enquirer* would be very high, but this is quite likely accounted for by the duplication as produced by the cell means, since the Sunday *Enquirer* audience was projected cell by cell from the audience of the daily *Enquirer*. In those cells in which the daily *Enquirer* has a high audience, the Sunday *Enquirer* will also have a high audience and a relatively high duplication is therefore assured with a random within-cell probability assignment. For the case of the other media, the random assignment of Sunday *Enquirer* probabilities of exposure will result in a rather high duplication because of the fact that the Sunday *Enquirer* reaches, by our best estimate, about 65 per cent of the adult population on an average issue. This seems already to insure a high duplication and therefore, no other adjustments will be made.

Duplication Among the Radio Audiences

The data relating to audience duplication for radio is very poor. One of the sources of data is derived from the Nielsen Radio Index as cited in Sandage.[5] These data indicate the proportion of the total audience of a program which listens throughout 95 per cent of the program. For the evening, this proportion ranges from 29 per cent for a band concert to 76 per cent for a leading variety

[5] C. H. Sandage, *Radio Advertising*, p. 147.

show. For the daytime, the percentages range from 50 per cent for a low-rated serial to 78 per cent for a high-rated serial. For two daytime news programs, 54 and 69 per cent were rather constant continuing listeners to the programs. One other piece of data cited by Nielsen indicates that the duplicated audience of two programs on different networks for the same evening was approximately 30 per cent of the audience of the smaller program. Also for two consecutive evening programs, the audience duplication was 71 per cent of the audience of the smaller programs.[6]

These data, however, are not exactly what we need since the percentages indicate the proportion of people who stay with one program or with both of two programs; in the case of our radio news broadcast vehicles over all of the five radio stations, we would like to know the number of people who, given that they listen at one point in time, are also listening one hour later. These people generally will be exposed then to news separated, on the average, by a one hour time interval. We do not wish to take account of switching from station to station since in general the news broadcasts occur on the hour over every station. We want the constancy of listenership over hours, not over stations. Therefore, the percentages

---

[6] A. C. Nielsen, New Facts, p. 55.

quoted from Sandage underrate the constancy of listening to the kinds of broadcasts, i.e., generic radio newscasts, about which we are concerned.

Since the data are so poor, we make an arbitrary decision as to how much duplication exists between audiences of two consecutive programs. We estimate that the duplication between any two programs is 50 per cent of the smaller of the two audiences. Using the smaller of the two audiences as the base takes account of the influx into the audiences of people arising in the morning and also the influx of people (men) returning from work during the hours from five to seven. The rationalization is that those who are in the audience at 7:00 A.M. are rather likely to be in the audience at 8:00 A.M. and those who are in the audience at 5:00 P.M. are rather likely to be in the audience at 6:00 P.M. This somewhat arbitrary formula defines the audience duplication for the radio audiences. As we noted above, we will not attempt any estimate of audience duplication between radio and newspaper audiences and we note that our only estimate of audience duplication for radio occurs between consecutive radio broadcasts.

Let us also note here the implications for duplication of our treating of weekday radio broadcasts as one group of broadcasts which apply to any weekday. This means, for example, that we do not actually have

five assignments of probabilities for the 8:00 A.M. news Monday to Friday, but that one assignment of probabilities holds for each of three days. Since the process is a probability process, this does not mean that a person who is exposed on one weekday will necessarily be exposed on another weekday, (given the proper messages) unless the probabilities are identically 1.0, which is a very rare case. However, the distribution of exposures within the cell will not be changing from weekday to weekday, but will remain rather constant. We believe that this adds a considerable amount of structure to the simulation. Since peoples' weekday habits tend to be rather constant, this added structure appears plausible.

The audience duplications between pairs of consecutive radio broadcasts are given in Table VI-2.

Upon completion of the fifth link of the simulation, the media system has been completely defined. We may conceive of the media system in the following schematic way. We form a matrix in which the entries in the left hand side are the identification numbers for each of the hypothetical members of the population. In the first columns of the matrix are entered for each member of the population the cell number or equivalently, the category or level for each of the population dimensions. In the next columns are entered, for each vehicle included in the media system, the individual's probability of

exposure to that vehicle and the distribution from which this probability was drawn. The schematic diagram is shown in Figure VI-2 below.

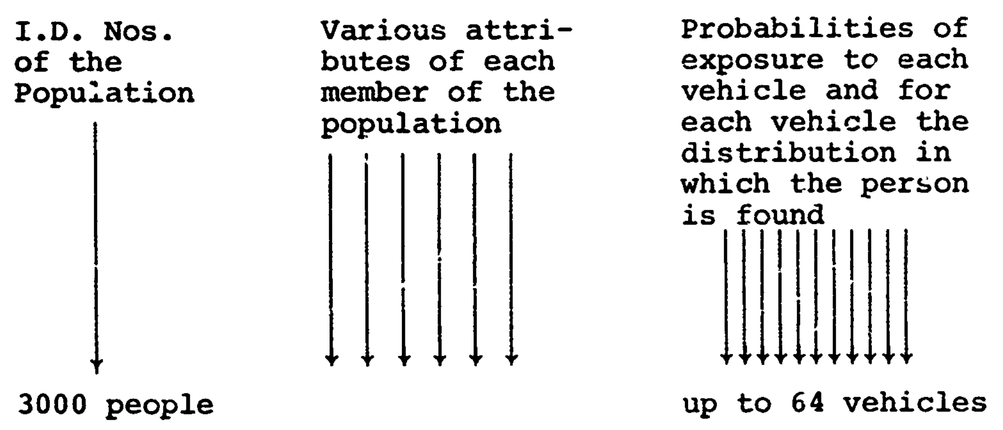


Figure VI-2. Schematic Diagram of the Media System and Population Upon Completion of the First Five Links of the Simulation

Table VI-2. Radio Audience Duplication for Pairs of Consecutive Radio News Broadcasts

Weekdays		Saturday		Sunday	
News Broadcast Pair	Audience Duplication (Per-cent of total pop.)	News Broadcast Pair	Audience Duplication (Per-cent of total pop.)	News Broadcast Pair	Audience Duplication (Per-cent of total pop.)
7-8 A.M.	3.17%	7-8 A.M.	4.93%	8-9 A.M.	1.96%
8-9	2.56	8-9	4.04	9-10	1.71
9-10	1.57	9-10	4.04	10-11	1.71
10-11	1.57	10-11	4.04	11-12	4.38
11-12	4.77	11-12	4.93		
12 A.M.-1 P.M.	5.00	12A.M.-1P.M.	5.40	12A.M.-1P.M.	7.89
		1-2P.M.	1.34	1-2 P.M.	1.26
3-4 P.M.	2.53	2-3	1.34	2-3	1.26
4-5	3.17	3-4	5.74	3-4	1.61
5-6	4.12	4-5	7.49	4-5	1.61
6-7	12.17	5-6	8.19	5-6	2.05
7-8	9.49	6-7	13.90	6-7	1.18
8-9	3.14	7-8	6.63	7-8	1.18
		10-11	1.27	8-9	1.28
11-12	1.76	11-12	1.55	11-12	2.08

CHAPTER VII

THE CONTENT ANALYSIS AND PREPARATION OF MESSAGE SCENARIOS

The final two links of the simulation introduce the actual scenario of messages ordered by themes and time periods and cycle through the population, testing to see whether each individual is exposed to the messages and reporting for each theme, time period, and media type, such statistics as the number of exposures and cumulative number of individuals exposed by media type.

The first step in exposing the members of the hypothetical population to messages appearing in the media system is the identification of kinds of messages of interest to the researcher followed by a content analysis of the media system, in order to locate the particular messages in the vehicle at a particular point of time and with a given format. Therefore, we turn now to the content analysis of the Cincinnati mass-media system during the six-month interval from September, 1947, to March, 1948.

<u>Themes Coded for Content Analysis</u>

In the literature reporting on the Cincinnati campaign, the major theme has to do with information about the United Nations. This is only natural, since the funding for the campaign and the purpose of the campaign were explicitly to change levels of information and opinions about the United Nations. For this reason, we initially began the study and simulation with the hope of using the panel data on attitudes about the United Nations as the major validation for the simulation. However, when we look at the results of the surveys, we find that, in fact, there was very little change in information or attitudes about the United Nations in the six month period. For example, the proportion of the population unfamiliar with the main purpose of the United Nations organization decreased from 30 percent in September, 1947, to 28 percent in March, 1948. The proportion of people who had heard or read anything about the veto power in the United Nations increased by only 3 percent from 34 percent in September to 37 percent in March. Although the level of information about the United Nations (as measured by the NORC questions) had not changed significantly during the six months, the opinions expressed about the United Nations did change somewhat. For instance, dissatisfaction with the United Nations increased from 28

percent of the sample in September to 33 percent of the sample in March. Similarly, among those queried, 62 percent felt in September that the United Nations would succeed while only 48 percent felt this way in March.[1]

Therefore, there were some net changes in opinions about the United Nations; moreover, if we look at the turnover within various subgroups of the population, the changes will be of differing magnitudes in different groups and some changes will be larger than those shown in the population as a whole. Therefore, in general, we will be able to correlate, or attempt to correlate, frequencies of exposures in subgroups of the population to various themes with the turnover or changes in opinion during the six month period. However, it should be noted that these changes are not the largest changes.

By far the most important changes in opinion during the six months had to do with the kinds of problems which were facing the world and the nation during this time. The main concerns were problems of peace or war and the relations of the United States with Russia. For example, when the sample was asked to think of the problem facing the United States, 24 percent in September and 46 percent in March first mentioned the possibility

---

[1] These data are taken from NORC Report No. 37A, Cincinnati Looks Again. The percentages are of those who have some minimal knowledge of the U.N. People who knew nothing about the U.N. were not asked these questions.

of another war or the problems of maintaining the peace. Sixteen percent in September and 29 percent in March first mentioned relations with Russia. Only 1 percent in September and 2 percent in March mentioned the United Nations in answer to this question. Another question asked about the kinds of problems in which the respondent takes a keen interest. In September, 54 percent of those asked expressed a keen interest in our relations with Russia; in March, this was 68 percent. In September, 51 percent expressed a keen interest in the control of the H-bomb; in March, this was 56 percent. When asked if they expected the United States to fight in another war within the next ten years, in September 48 percent answered "yes," and in March 73 percent answered "yes". Thus, it seems clear that changes in public opinion on the issues of war and peace and relations with Russia and possibly on the issue of the control of the atomic bomb were much greater and more important than changes in opinion about the United Nations. Therefore, in addition to coding the newspaper stories in terms of themes related to the United Nations, we also coded these stories for items related to war, violence and threats to peace, to the United States' relations with Russia, and to control of the atomic bomb. In fact, it will be shown below that the largest number of items related to matters of war and peace.

The number of themes used in the content analysis totaled 17 including themes having to do with U.N. peacekeeping or the veto power in the United Nations, the U.N.

and human rights, etc. With these kinds of themes, it is obvious that a single news item might carry more than one theme, e.g., the story on the Palestine debate in the United Nations might emphasize both the peacekeeping role of the United Nations (Theme 1) and also the dissension and dispute among the great powers (Theme 2). In this case, the story would be recorded for both themes.

The themes chosen were, therefore, related to the questions asked in the NORC survey. It was felt that these themes, more than any of the others in the press at this time, would be likely to explain the changes in attitudes, information, and opinions during the six month period which were found in the survey. In the Appendix to this chapter, we present the themes in the content analysis, a description of each of the themes, instructions for the coders, the kinds of newspaper articles which might be included as relating to a theme and, for the benefit of the coders, a list of questionnaire items and the results from these items which might be thought to be closely related to each of these themes. Also a sample coding sheet is included. Since there exists no record of the content of radio news broadcasts for the period, the coding was limited to the four major newspapers; the kinds of messages broadcast will be inferred from the treatment in the newspapers.

Table VII-1 shows the results of the content analysis for the 17 themes. A total of 1,428 different stories

were coded (including the stories duplicated for the reliability estimation) resulting in a total of 1,584 occurrences of the 17 themes in the four newspapers. Of these, the most prominent theme by far was the theme entitled "Acts of Violence, Threats to Peace, and War." Other prominent themes included "U.N. Peacekeeping," "Discussions of Russian and American Relations," "Dissension Among the Great Powers in the U.N." and "Mention of the Organizations Sponsoring the Information Campaign." Given the relative number of occurrences of each of the themes, it is not surprising to find that fears of war and interest in Russian-American relations were among the areas showing the greatest change in the panel. With reference to the negligible change in information concerning the veto power, we note that there were very few mentions of the veto power in the press. (The category "Any Reference to the Veto Power," which one would assume would be larger than the following two categories which specify something particular about the veto power, actually is smaller than the other two categories. This probably reflects the fact that the coder had a much easier time in keeping in attention specific kinds of coding and would likely code one of the particular kinds of veto power. Nonetheless, even with this kind of inaccuracy, the mention of the veto power, or of other areas of concern to the U.N. were very rare in the press at this time.) Also, we note that there

Table VII-1. Content Analysis for Seventeen Themes Occurring in Cincinnati Newspapers, September 16, 1947 - March 7, 1948.

Themes Coded	No. of Newspaper Occurrences	Est. No. of Radio Occurrences	Total No. of Occurrences of the Theme
Themes Relating to the U.N.			
1. U.N. Peacekeeping	234	361	595
2. Great Power Dissension in the U.N.	108	249	357
3. Any Reference to the Veto Power	22	32	54
4. Veto Only in Security Council	27	63	90
5. Veto Only by Great Powers	28	84	112
6. U.N. and Human Rights	20	12	32
7. U.N. and World Trade	5	11	16
8. U.N. and World Health	9	7	16
9. UNESCO and Health	14	0	14
10. Slogan of the Information Campaign	1	0	1
11. Mention of the Information Campaign Sponsors	100	42	142
12. Explicit Explanation of the U.N.	31	12	43
13. Satisfaction with the U.N.	19	44	63
14. Dissatisfaction with the U.N.	17	58	75
Themes Relating to Other International Issues			
15. Control of the Atomic Bomb	91	129	220
16. Russian-American Relations	217	535	752
17. Violence, Threats to Peace, War	643 (514)[a]	1559 (1247)[a]	2202 (1761)[a]
Total	1584		

[a]The simulation allows for a maximum of 1800 occurrences of any single theme; therefore, one of every five of the occurrences of the last theme were omitted in constructing the scenario of messages. The values in parentheses are the reduced scenario values.

were very few expressions of satisfaction or dissatisfaction with the United Nations. One might assume that in this case the coders failed to categorize in this manner messages which implicitly carried notions of satisfaction or dissatisfaction with the U.N.; one might imagine here that discussions of great power dissension in the U.N. or reports of such, carry with them an implied attitude of dissatisfaction with the U.N. The coders, however, did not make this assumption. Finally, we note that there were a total of 31 occurrences in the newspaper of stories containing some explicit explanation of the purposes, workings, or functioning of the U.N. It seems quite obvious that although the mention of meetings of the sponsors of the campaign were often carried in the newspaper, the campaign did not succeed in significantly changing the number of stories in the newspaper which conveyed some explicit explanation of the United Nations. (Even had they been able to do so, these stories would not have been widely read, especially by those who had little knowledge.) This makes much less surprising the finding of very small changes in opinion and attitudes and knowledge about the U.N. over the six-month period. (Since no more than sixteen messages were coded for themes 7, 8, 9, or 10, these themes were not included in the simulation scenarios.)

The Cincinnati campaign may not have demonstrated the difficulty in reaching certain segments of the population with information about public affairs (although this

may be true enough); what it seems actually to have demonstrated is the difficulty in presenting information and explicit explanation about some international issue in the mass media. The campaign sponsors thought they had done this; however, the content analysis seems to indicate that they did not succeed nearly as well as they had imagined in getting the information into the mass media.

Because it is only possible to calculate duplication between any theme and the first three themes of the simulation, the order of the themes in the running of the simulation was changed to place the most important themes first. The actual order of themes in the scenario (and so presented in Chapter VIII analyzing the output statistics, except for the final theme which was lost due to a programming error) follows:

1. U.N. Peacekeeping
2. Russian-American Relations
3. Violence, Threats to Peace, War
4. Great Power Dissension in the U.N.
5. Any Reference to the Veto Power
6. Veto Only in Security Council
7. Veto Only by Great Powers
8. U.N. and Human Rights
9. Mention of the Information Campaign Sponsors
10. Explicit Explanation of the U.N.
11. Satisfaction with the U.N.
12. Dissatisfaction with the U.N.
13. Control of the Atomic Bomb

Coding Reliability

In the coding of the content analysis of the Cincinnati newspapers, two coders read the total issues of the three newspapers, the Enquirer (both daily and Sunday), The Times-Star, and the Post. In order to test the reliability of the analysis, each of the coders was required to code each of three one-week periods. These periods were the first week in November, the first week in December, and the first week in January. However, several thorny difficulties ari[illegible] in making a calculation of the reliability for this kind of content analysis. The first of these is that there is, unlike the case of coding survey responses, no known number of items which must be coded, i.e., the number of items relevant to any of the seventeen themes in the content analysis must be decided upon by the judgment of the coders and is not known beforehand. Therefore, we have decided to use as a base the total number of stories coded i[illegible]ese th[illegible] weeks by the two coders, realizing that some stories may have escaped the notice of both of the coders and so th[illegible] "N" may be larger than the "N" given below. However, it would seem that the most important and the most salient stories in terms of our seventeen themes would probably be caught by at least one of the two coders.

The total number of stories coded by the two coders in the three weeks in three newspapers was 240 stories. Of these stories, 19[illegible] 84 percent were coded by Coder No.1 and

149 or 62 percent of the stories were coded by Coder No. 2. The number of stories in common between the two coders was 102 or 42.5 per cent. Because Coder No. 1 seemed to find significantly more stories relevant to the themes than Coder No. 2, whenever both coders coded a newspaper, the coding of Coder No. 1 was used. Of the total number of issues of newspapers coded, Coder No. 1 coded 53 percent in the data used for our total content analysis. Also, in interviewing the Cincinnatians involved in the Cincinnati Plan, these people indicated that the Cincinnati Post was the most important newspaper in terms of dissemination of news about the U.N. and international affairs in Cincinnati. Therefore, Coder No. 1 was assigned to this newspaper and coded 100 percent of the issues of this paper. Coder No. 1 also coded 47 per cent of the Times-Star issues and only 12 per cent of the Enquirer issues. Considering the proportion of issues coded by Coder No. 1 and the relative percentages of the total number of stories which were coded by each of the interviewers, we estimate that about 75 percent of the stories relevant to our seventeen themes were coded by the combined coding operation. Therefore, 25 percent of the relevant stories were probably not coded and were not included in the scenarios. We would guess that these stories which were missed were, however, less important and less obviously salient to the themes than the stories which were coded. Nevertheless, the reader may keep in mind the fact

that perhaps 25 percent of the stories that have some sort of relevance to each of these themes were not, in fact, included in the coding or the scenarios.

A second major problem in evaluating the reliability of the coding operation is the fact that the categories into which each story was coded were not mutually exclusive, i.e., a story might very well contain several of the themes of the content analysis. For example, it would not be uncommon for a single story to contain both the themes of the U.N. peacekeeping and of dissatisfaction with the U.N. The formula recommended for estimating the reliability of a coding operation is Scott's formula for Pi, which measures the ratio of the actual difference between the observed and expected agreements between coders to the maximum difference between the observed and expected agreement.[2] Thus, Scott's Pi is a measure of coder reliability which varies between the limits of 0.0 and 1.0. However, Scott's Pi assumes that each of the coding operations produces one and only one coding assignment for each story. If we disregard this assumption, we are essentially throwing out the fact that not coding a story as relevant to a particular theme by

---

[2] William A. Scott, "Reliability of Content Analysis; a Case of Nominal Scale Coding," Public Opinion Quarterly, (Fall 1955), 321-25.

both coders is an instance of agreement. At any rate, disregarding this fact, the value for Scott's Pi is .54 indicating only a low degree of reliability in the coding of the Cincinnati papers.

There is another way to look at the reliability of the coding, a way which has faults but may nevertheless be instructive to the reader. We might look at the total number of themes which were assigned to the total number of stories which were coded in common and the number of these which were in agreement. (This measure is biased by the fact that if we have only two possible themes (i.e., coding categories) we have a high probability of agreement, and as we have more, i.e., seventeen themes to assign, we have much less possibility of agreement simply on the basis of chance.) This measure shows 63.3 percent of the coding choices in agreement. We might also look at the individual themes for the proportion of coding in agreement. The most important themes (in terms of number of stories) were Themes Nos. 1, 2, 11, 15, 16, and 17. For these themes, the proportion of the total choices of the theme which were made by both of the coders are: Theme 1 - 50 percent, Theme 2 - 33 percent, Theme 11 - 89 percent, Theme 15 - 89 percent, Theme 16 - 40 percent, and Theme 17 (which was by far the most common theme) - 89 percent. Judging from the kinds of information contained in each of these themes, it seems clear that the coders had a high degree of agreement about such things as threats to peace or

wars (violence) but for such other less clear-cut topics as dissatisfaction with the U.N. or the U.N. and world health, the agreement was much lower.

We may summarize our findings about the reliability of the content analysis in this manner; there was relatively high reliability in coding the stories relevant to one or another theme in the content analysis. For certain themes, the agreement between the coders on the relevance of the story for the theme was very high. These themes were Themes Nos. 11, 15, and 17, which are the "Campaign Sponsors for the U.N.," "Control of the Atomic Bomb," and "Threats to Peace, Wars." The content analysis could surely have been improved with more and better training of the analysts. One of the obvious problems was that Coder No. 2 did not attend to stories of possibly lesser significance for the seventeen themes.

While there is room for improvement in this content analysis, it nevertheless seems reasonably accurate as input for the simulation scenarios, especially in light of the level of precision in specifying the various parameters of the media system, and the kinds of assumptions necessary to generate the radio messages.

Estimation of the Radio Messages

The second column in Table VII-1 contains the number of radio occurrences of each of the themes. There was no way to directly ascertain what kind of messages were broadcast

during the six-month period of their distribution throughout the media system. Therefore, some estimate had to be made on the basis of the kinds of stories that were carried in the newspapers. This was done in the following manner; first, it was assumed that there was a correspondence between the messages appearing in the morning and early afternoon newspapers, i.e., in the daily Enquirer, and the radio news during the morning and early afternoon. Likewise, we assume that there was a correspondence between the news in the two afternoon newspapers, the Times-Star and the Post, and the afternoon and evening radio news broadcasts. Finally, we assume that the Sunday news broadcasts reflected the news in the Sunday Enquirer. However, the amount of time devoted to international news in news broadcasts is quite limited and not nearly all of the stories or themes found in the newspaper would, in fact, be found in the corresponding radio broadcast. Therefore, we used the ad hoc formula that only if a story appeared on the front page of the newspaper would it appear in the corresponding radio broadcast. On this basis, we estimated the occurrence of the themes in the radio broadcasts and these are the numbers which are presented in the second column of Table VII-1. It will be noted that in general these numbers are much larger than the number of occurrences in the newspaper even though they reflect only those newspaper occurrences which are found on the front page. This is because in general one front-page newspaper

newspaper occurrence was likely to generate from five to ten radio occurrences, one at each hourly time slot when the news broadcast took place. Thus, the mention of one of the themes on the front page of the daily Enquirer would imply that the theme was broadcast in the 7:00 A.M., 8:00 A.M., 9:00 A.M., 10:00 A.M., 11:00 A.M., 12:00 A.M., 1:00 P.M., and 2:00 P.M. news broadcasts. It is true in some cases that the number of radio occurrences is less than the number of newspaper occurrences or nearly equal to the number of newspaper occurrences. This is an indication that most of the newspaper occurrences did not take place on the front page of the newspaper; that is, the stories were not considered most important news. These stories are generally just what we would expect, i.e., themes relating to the U.N. and human rights, the U.N. and world trade or world health, UNESCO, slogan of the information campaign, explicit explanation of the U.N., or references to the veto power. Thus, the relative increase in number of radio occurrences over the number of newspaper occurrences can be taken as an indication of the relative importance of each occurrence of the theme, i.e., its location on the front page.

There is one other note of explanation about the table. In the simulation, a maximum of 1,800 occurrences of any single theme is allowed; therefore, it was necessary to somehow cut down the number of stories or occurrences of the final theme which totaled 2,202 occurrences combining

both newspaper and radio. We did this in a rather arbitrary fashion, simply discarding one out of every five of the occurrences of the last theme. Although at this stage of the simulation, the effect of this paring down of the number of messages is not obvious, we shall find upon examination of several features of the output statistics (in Chapter VIII), that most probably the effect is to diminish the overall average level of exposure, but not to change significantly the distribution of exposure over the population types. The figures resulting from this paring, plus the figures for the other themes, are the basis for the scenario inputs to the exposure part of the simulation.

It is possible that there may be some exaggeration in the number of radio stories created by this estimation procedure because of the possibility of the same story with identical themes being carried on the front pages of both of the evening newspapers. In this case, the story would appear twice in each of the radio news broadcasts corresponding to the afternoon newspapers. This kind of duplication appears unavoidable since it would be very complex to detect from the coding whether the exact story in the same form was presented in both of the evening newspapers. In order to check the amount of error possible here, we have calculated the total number of front page stories appearing in the _Times-Star_ and the _Post_. In the _Times-Star_, this total is 50 and for the _Post_ the total is 128. Therefore, the largest number of stories which could

possibly be erroneously duplicated in the radio news broadcasts is 50 and more than likely it is significantly smaller than 50. (The figure 50 would be true if each of the 50 stories that appeared on the Times-Star front page also appeared on the Post front page carrying the same themes.) Therefore, we conclude that there is a likelihood of some exaggeration in the number of evening radio occurrences of the themes, but that it is relatively small. Given the underestimation of the total number of occurrences of themes, in the newspaper (we estimated above that about 75 percent of the stories which could have been coded were actually coded) this error is even less important.

The Estimation of Newspaper Message Exposure Probabilities

The simulation up to this point has produced probabilities of exposure to the various vehicles in the mass media system. If we were to run a simulation based on these probabilities, the resulting exposures would be exposures to the vehicle. However, what we want is the exposure of people to messages contained in the vehicles. For this exposure we need some estimate of the probability of exposure of an individual to a message in a vehicle, given that the individual has been exposed to the vehicle. Then the probability that the individual is exposed to a particular message in a vehicle is just the product of the two probabilities. Let us call the probability of exposure to a message in a vehicle, given exposure to the vehicle, a message exposure probability.

Exposure to the message depends upon two factors; the interest of the reader in the topic and the play which the message gets in the format of the medium. If for each theme in the simulation, we had a large number of stories of differing formats, with an estimate of the proportion of the vehicle audience who were exposed to each story, then we could, in principle, separate out the effects of formating from the effects of the content of the theme upon the message exposure probabilities. Unfortunately, we have not been able to find a large sample of news stories for each of the themes and for the 1947-1948 time period for which these exposure data are available. We have found 127 stories relating to general international news for which these data are available; from them we have estimated (using multiple regression) the effects of format on the conditional message exposure probabilities for international news.

The basis for the estimation of message exposure probabilities is a study done by the Advertising Research Foundation called The Continuing Study of Newspaper Reading which is actually a series of 138 studies of different newspapers dating from 1939 to 1950. The aim of this series of studies was to look at the readership of various kinds of articles in newspapers given that a person had read at all in the newspaper. We have used the data on international news stories in the 1947-1948 studies for a multiple regression predicting readership of men and women based on various format factors.

Each of the studies was conducted in the following manner. A quota sample was selected representing various economic levels in the city zone of the newspaper in their proper proportion and paralleling (geographically) the circulation of the newspaper. This sample consisted of 225 men and 225 women who were admitted readers of the previous day's issue of the newspaper. When the interviewer contacted a respondent according to this quota sampling procedure, and the respondent had read at all in the previous day's issue of the paper, the interviewer then presented the respondent with a copy of that issue of the paper and asked the respondent to go through the newspaper page by page marking with crayon the articles, pictures, and advertisements which he remembered having read the previous day. The interviewers were instructed to use only two questions in their interviewing. The first was, "Did you happen to see or read anything on this page?" and the second, "Did you happen to see or read any of the advertisements on this page?"[3]

The data from these 450 interviews is presented in an exact copy of the issue of the newspaper, printed

---

[3] A summary of the studies, the findings, and the methodology is presented in <u>The Continuing Study of Newspaper Reading: 138 - Study Summary</u>, The Advertising Research Foundation (New York: Advertising Research Foundation, Inc., 1951).

in a large journal form. Superimposed on each article, advertisement, picture, or cartoon is the proportion of men and the proportion of women who reported having read the article. We have then, for each article in the newspaper, the proportion of men and the proportion of women who admit to having read the article, given that they have been exposed to the vehicle at all. This is the basis for our estimates of message exposure probabilities. (Since exposure data are reported only by sex, this is the only demographic variable considered in calculating message exposure probabilities.)

Of the 138 studies, Studies No. 110 through 120 occurred during the time period from June 19, 1947, to June 3, 1948. This is a total of eleven studies of which nine are evening papers and two are morning papers. A list of these papers, the dates, study numbers, and circulations, is given below. As can be seen from the table, the newspapers range in size from the Athens Messenger, (Athens, Ohio) with a circulation of 18,805 to the Washington Post (Washington, D.C.) with a circulation of 165,554.

From these 11 newspapers a total of 127 stories dealing with international political news[4] were coded for

[4] Typical titles for these stories included: "Vandenberg Plan Appears Assured of U.S. Blessing," "Russian Soldiers Outnumber Yanks by 20 To 1 Ratio," "My Day" by Eleanor Roosevelt, and "Pravda Assails Acheson's Speech as Rude Slander."

Table VII-2. Newspaper Studies from Which International News Items were Coded for Prediction of Exposure Probabilities

Study No.	Newspaper		Date	Circulation[a]	No. Items Coded
110.	Easton Express	E	6/19/47	41,424	16
111.	Danville Commercial-News	E	7/24/47	33,073	5
112.	The Washington Post	M	9/25/47	165,554	21
113.	Waterloo Daily Courier	E	10/23/47	46,388	5
114.	The Birmingham News	E	11/20/47	151,001	6
115.	Stamford Advocate	E	12/11/47	19,188	14
116.	Hollywood Citizen-News	E	1/28/48	38,363	12
117.	The San Francisco News	E	2/4/48	143,058	16
118.	The Battle Creek Enquirer and News	E	3/31/48	31,257	13
119.	The Burlington Free Press	M	4/28/48	23,947	14
120.	The Athens Messenger	E	6/3/58	18,805	5
	TOTAL				127

M - Morning paper; E - Evening; D - All-day; S - Sunday

[a]The latest publisher's semi-annual A.B.C. statement available at time of publication of that study.

several format characteristics and the percentages of men and the percentages of women exposed. One set of variables used to describe the story formats referred to the page location of the story. The categories were (1) front page lead story, (2) other front page, (3) second or third page, (4) editorial page or pages, (5) women's page or pages, (6) other inside pages. Other variables described whether the story was in print or a picture or both, whether it was located at the top of the page, the middle, or the bottom of the page, the caption width in columns, and the number of column inches in length. (Of course, for all except the last two of these variables the coding was a 1 or a 0 as the variables are actually dummy variables.) Then the linear regression was run in a stepwise fashion, independent variables being added one at a time in the order of their contribution to the explanation of the variance in the dependent variable. In this way the most important of the independent variables were isolated.

Results of the Multiple Regression

First we examine the mean exposure probabilities for men and for women for the 127 stories. For the men the average exposure was 21.6 percent with a standard deviation of 13.49 percent. For the women the mean exposure was 13.0 percent with a standard deviation of 11.12 percent. These numbers are validated somewhat by

a study done by Swanson which is a summary of all 138 studies in the Continuing Study.[5] Swanson found the following summary readership figures for various categories of articles. For war news, the mean male readership was 39.4 percent and the mean female readership was 29.8 percent; for defense news, males 30.8 percent, females 27.4 percent; for economic-social international relations news, male readership 22.1 percent, female readership 18.1 percent; for political international relations news, male readership 24.0 percent, female readership 15.5 percent; for general politics news, male readership 19.6 percent, female readership 14.0 percent. The differences between Swanson's figures, which are somewhat higher than the means in the 127 articles, probably come from changes over time and the categorization of the articles. At any rate, Swanson's figures for general international news are not widely different from the figures for the stories recorded during the year from June 1947 to June 1948.

In the table below we present the result of the stepwise linear regression. For men we have used a regression equation with six variables and a constant which has a multiple R correlation coefficient of .6900

---

[5]Charles E. Swanson, "What they Read in 130 Daily Newspapers," Journalism Quarterly, 32:4, (Fall, 1955), 411-421.

Table VII-3. Regression Coefficients and Significance Levels for Predicting Message Exposure Probabilities from Format Variables for Men and Women

Format Variables	Regression for Men		Regression for Women	
	Coefficient	Significance Level	Coefficient	Significance Level
Print	-24.8685	.0000+	-21.1265	.0000+
Front Page Lead Story	18.1688	.0003	8.6475	.0006
Number of Columns Width	.0274	.0024	.0151	.0079
Other Front Page	10.5324	.0013	--	--
Editorial Page(s)	8.3395	.0016	--	--
Bottom Third of Page	- 3.8694	.1757	--	--
Other Inside Pages	--	--	- 5.3705	.0161
Picture	--	--	4.7442	.0665
Women's Pages	--	--	10.8870	.0513
Second or Third Pages	--	--	- 6.1083	.0021
Constant Term	35.3341	--	31.4784	--
R^2 (multiple)	.476	.0000+	.528	.0000+
Standard Error of Estimate	9.7619	--	7.6384	--

(significant at the .0000+ level) and which indicates that the format factors explain about 48 per cent of the variance in the dependent variable, the proportion of men reading the article. For women seven independent variables were included which together produce a multiple R of .7267 (which is significant at the .0000+ level of probability) and which account for approximately 52 per cent of the variance in the female readership. We note also that there are several rather plausible differences between the men and women in those independent variables which contribute most to the prediction of readership. The location of the article on either the front page or the editorial pages is important in prediction of male readership but not for female readership; however, the location on the second or third pages, other inside pages, on women's pages, or the presence of a picture is important in prediction of female readership but not for male readership. These equations were then used to predict the readership of the stories actually coded in the content analysis of the Cincinnati press on the basis of the formating of the stories.

At this point perhaps it is appropriate to comment on the significance of these equations. It seems reasonable, especially given the fact that there are no other methods available, to use these equations to predict the readership for men and women of the Cincinnati news stories. The equations were derived from data which is in many respects similar to the Cincinnati news stories

both in the content and in the time of the story. Although we cannot be sure that the population of news stories of which our 127 are a sample is equivalent to the population of news stories in the Cincinnati newspapers, for present purposes we are willing to make this assumption.

Another factor is quite obvious from the statistics of the regression. The format factors, at least as they are characterized in the present content analysis, serve to explain only half of the variance in the readership figures for men and women. We might ask, then, what additional variables would be necessary to more completely specify this readership. In reading and coding these stories, it became obvious to the author that certain phrases or words in the caption of the stories serve to attract the attention of the eye and increase the likelihood of reading the story. For the author (a male) words such as "seize," "threaten," or phrases such as "Russian Soldiers Outnumber," or "Bigger Atom Bombs" served to attract attention. We theorize then that much of the remaining variance in the readership of these stories is dependent upon the effects of certain key words or phrases in the captions upon the audience. One might try to design an experiment which would try to separate the effects of these kinds of phrases from the formating, etc., but at first glance this seems to be a very difficult research problem. There is, however, another way to attack the problem. One might imagine that the editors or writers who set the captions for the stories are

very astute at picking phrases and words which, given the content of the story, will best attract the readership of the audience. Thus one might be willing to make the assumption that the content of the story will be represented in the most favorable way in the caption for the story and that a description of the effect of the content of the story itself would go a long way toward explaining the effects of certain of these key words. This is an area for further research and we have not seen any research on this kind of effect in any of the journals.[6] At any rate, it is obvious that even within the limited message contents

---

[6] The key words or phrases in the caption obviously have an effect upon the readership of the article but also the content of the article has an effect upon the words, phrases, that appear in the caption. One possible model might assume an average level of attention due to the content of the article and perhaps a more specific additive effect upon attention if the content of the article allows the use of certain kinds of words or phrases which are superior attention-getting words and phrases. Also it would seem likely that the attention paid to a given word or phrase by the reader depends upon the characteristics of the reader, in particular upon his sex, education, and quite likely some dimensions of personality.

In much the same vein, Tannenbaum and Lynch differentiate between two dimensions in their conception of sensationalism, i.e., topical and stylistic sensationalism. They present a factor analysis of twenty semantic differential scales which identifies the evaluative, the excitement, and the activity factors underlying the concept. However, they make no attempt to isolate the effects of single words or phrases, or to determine if interaction exists with personality variables in the reader. See Percy H. Tannenbaum and Mervin D. Lynch, "Sensationalism: The Concept And Its Measurement," Journalism Quarterly, 37:3 (1960), p. 381-391.

and themes of the present analysis there seems to be enough variety of content or theme and variety of attention gathering expressions and phrases in the captions of the messages so as to limit to about 50 percent the amount of variance in readership which can be explained on the basis of format factors alone.[7]

Generation of Exposure Probabilities for Occurrences of the Themes in Radio News Broadcasts

For any occurrence of a theme in the newspaper, we are able to generate a probability of exposure based on the multiple regression developed above. Every newspaper message is assigned a probability of exposure for men and for women from this procedure. The next problem is to estimate these probabilities of exposure for the radio occurrences of these messages. We did this by making some assumptions about consistency between the appearance of the stories in the newspaper and the length and emphasis they would be given in a news broadcast. We assumed that the newspapers and the radio make consistent news judgments when we decided to attribute to the radio broadcasts those messages which occurred on the front page

---

[7] Clinton R. Bush in a research note "Content and 'Mise en Valeur': Attention as Effect," Journalism Quarterly, 37:3, (1960), p. 435-6, presents a formula for weighting journalistic text depending on elements of format and typography developed by a French professor, Jacques Kayser. Bush opines however, that "...the words in the headline are probably more important than the display."

of the newspapers. At this point, we go one step further and assume that those front page stories which had the highest probabilities of exposure in the newspaper will also be given the kind of emphasis that will result in their having the highest probabilities of exposures in the radio broadcasts. However, it is difficult to imagine that the probabilities of exposure are equal for stories in the two media. It seems reasonable to assume that, even given that the person is exposed to the newspaper as a vehicle, he has a great amount of choice in the actual story to which he attends. Therefore, we expect quite a difference between the number of people exposed to the vehicle and the number of people exposed to any given message in the vehicle, i.e., we assume that the message exposure probabilities for a given newspaper message will be rather low and indeed we have found them to have means of 21 and 13 percent for men and women respectively. However, we assume that the possibility for choice in attention to the various radio messages is not so strong as with newspapers, given that the person is exposed at all to the broadcast. It will be recalled that the definition of a member of the audience of a radio vehicle is someone who could answer certain questions about the time, station, and content of the vehicle. The probabilities thus generated refer not to the instantaneous listenership of the vehicle, but the average listenership, i.e., the probabilities represent people who have listened to the entire program or have at least tuned in for the duration of the program. Also we have estimated (p. 121) that international news occupied about three minutes of a fifteen-minute newscast.

Thus, the probability that a listener is exposed to any given message in the vehicle is likely to be higher than the probability of a newspaper reader being exposed to a given message. There are some data which shows that these probabilities are indeed somewhat higher. Studies of housewife viewers of television programs have shown that about 25 percent of these viewers can recall the commercial of the program 24 hours later. Contrast this to the probability of exposure to a given message in the newspaper which for women was 13 percent.[8] Of course, some of this difference may be accounted for by the fact that these commercials are the kind of information that are aimed at women and presumably women have a higher likelihood of attending to these sorts of commercials rather than the average story in a newspaper. Nevertheless, it offers some support for the fact that the 24 hour recall of stories in electronic vehicles by a member of the average audience is somewhat larger than the 24 hour recall of the particular message in a newspaper. Therefore, in calculating the probabilities of exposure to the radio messages, we have made a systematic increase in the probabilities generated for the newspaper occurrences.

---

[8] These data are taken from studies by Foote, Cone, and Belding in Queens and St. Louis, which found twenty-four hour recall of 26 percent and 23 percent, and Look Studies of Advertising Communication which found 24 percent recall after 24 hours. The studies are cited in a volume called Media Balance, Life Incorporated, 1965, p. 7.

Obviously, the probability 1.0 should not be increased and probabilities near this value should be only slightly increased, but in such a way so as never to exceed the value 1.0. We have also decided that the value 0.0 should remain the same and that the maximum increases should occur for probabilities of 0.5. With P' representing the transformed values, the following parabolic function effectively accounts for the ceiling and floor effects:

$$P' = -0.8P^2 + 1.8\,P.$$

Probabilities equal to zero remain at zero. Probabilities equal to 1.0 remain at that value, but probabilities intermediate between those are increased with the maximum increase happening to a probability of .5 which is increased to .72. The table below shows the increases in probabilities for various values of the newspaper estimated probabilities. If we assume that most of the newspaper probabilities were on the order of .1, then the transformation causes an increase in probabilities of about 70 percent. However, over the entire range the increase is 13 percent.

Table VII-4. Typical Values of Transformed Probabilities

Original (Newspaper) Probabilities	Transformed (Radio) Probabilities
0.0	0.000
0.1	0.172
0.3	0.468
0.5	0.720
0.7	0.868
0.9	0.972
1.0	1.000

Some Observations About the Message Exposure Estimates

We note here that the definition and data used in generating both the newspaper and radio probabilities refer to the chance that a member of the audience of a vehicle will be exposed to and recall a message in the vehicle 24 hours after the appearance of the vehicle. Therefore, the exposures which are output from the simulation are really estimates of (aided) recall 24 hours after the appearance of the vehicle. When we cumulate these sorts of exposures over a significant length of time such as six months, we are obviously making some distortion in the kind of knowledge which is imparted by exposure to these messages. It seems quite likely that there is some decline over a six month period in recall of the message from the proportion recalling at the end of 24 hours. However, on the other hand we are not in general trying to ascertain whether a person has been exposed to a particular message but rather

some general effects of exposure to messages. This is the kind of thing that is measured by the questions asked in the NORC survey. One might imagine that these general and perhaps unconscious or unrealized effects diminish with time or perhaps even increase (i.e., "sleeper effect") in time much less rapidly than recall of exposure to a given message. We suggest, therefore, that we may hope to find a correlation between the kinds of people frequently exposed to a certain theme and questions of information or attitude at the end of a six month period which are related to that theme.

Finally, it should be noted that the simulation takes these probabilities for message exposure for men and women and uses them in a nonobvious way in the simulation. In the actual simulation we have made assumptions about the relative probabilities of exposure to the message given exposure to the vehicle for breakdowns of the audience other than simple sex breakdowns. Thus, from our two probabilities we make other assumptions which generate a set of 36 probabilities which distinguish between 36 audience types (each population dimension excepting interest is included) but which nevertheless preserve the probabilities for men and for women for each of these messages that we have calculated up to this point.

There is one obvious failing in this procedure, however, in that these mean message exposure probabilities for a particular audience type are applied to each person

in the audience type. This is just the kind of situation which we went to some lengths to avoid in the first pass of the simulation where we use cumulation data to distribute individual probabilities about the audience type mean in order to produce the cumulative patterns of exposure which are known to exist. In order to do this at this stage of the simulation would require a tremendous amount of additional programming, computer space, and time, and for this reason this has not been done. We have, as a matter of fact, no data which would estimate anything like cumulation in this case and so we cannot go any further, but we note that this is obviously a simplifying assumption which is not strictly within the philosophy of the previous links of the simulation.[9]

---

[9] One possibility for generating probabilities different from the mean for each group would be to take the mean plus another figure drawn out of the air which would produce a beta function distribution of the probabilities for the group about the mean. This other figure drawn out of the air would be chosen in such a way as to make this beta function a bell-shaped curve about the group mean. Only the parameters for this beta function would need to be stored in the computer and as a member of an audience type was processed, a probability would be drawn at random from the appropriate beta function. Without knowing the cumulation, we could not produce it in the choice of the beta function but the procedure would doubtless generate a more realistic set of probabilities than the procedure of simply using the mean probability for every member of the audience type.

<u>The Simulation Model of Message Exposure</u>

Consider what might be a reasonably complete description of message exposure probabilities for the simulation. How shall we describe the message exposure probabilities for a single message in the scenario? We might, on the basis of the theme of the message and the format factors, generate a set of probabilities of exposure to the particular message. These probabilities might be as detailed as mean probabilities for every audience type in the simulation. This would mean that we could have up to five hundred mean probabilities of exposure associated with each message in the scenario. Thus, the probabilities of exposure associated with any one message may be functions of the content of the message, the formating of the message, the vehicle in which the message is carried, and finally the audience type for which the projection is to be made. Clearly, description in this much detail for each message event in the scenario is quite beyond storage or time limitations of the simulation. Therefore, it was necessary to devise a simplified model, throwing out some of the kinds of interactions which might be possible, in order to arrive at some parsimonious description of the message exposure probabilities.

Moreover, in designing the model we wanted to take account of the fact that, in general, the researcher will not have detailed information about the message

exposure probabilities for each message, but will know only an "average" conditional probability of exposure to a theme contained in a vehicle. (One often encounters statements like "About twenty percent of newspaper readers see the average foreign affairs story.") Also, the researcher will not in general know the absolute value of the message exposure probabilities for various groups, but will be able to make statements like, "Given readership at all in the vehicle, men are one-third more likely than women to read international affairs news." The following simplified model assumes only this minimal knowledge: we consider a group of vehicles and audience types for which there is some average proportion of the vehicle audience who are generally exposed to a given theme. This is an average audience of the theme, "averaged" over all possible relevant messages and formats. For this theme and group of vehicles we define, for any dimension describing the vehicle audience, ratios between the probabilities of exposure to the message, given exposure to the vehicle, for any two adjacent levels of the dimension.

We assume that these ratios are constant for all messages in this theme appearing in this group of vehicles. For example, the <u>ratio</u> of the message exposure probability for women to that for men is assumed a constant over all messages and all vehicles in the vehicle group. (We have not assumed that the <u>probabilities</u> of exposure for men or women are constant for all vehicles in a vehicle group

because the vehicle audience composition for the various vehicles may differ to quite a degree. If a constant proportion of that vehicle audience sees the theme generally, we cannot then assume that the proportion of men and proportion of women are also constant, since the composition of the vehicle audience (the number of men and women) changes from vehicle to vehicle.[10]

These assumptions lead to the following equations: let (V_{Mi}) and (V_{Wi}) represent the number of males and females, respectively, in the i^{th} vehicle audience. These figures (or the breakdown by any other population dimension) are known from the previous links of the simulation. Let (A) be the proportion of the vehicle audience who generally sees the theme. This is a known "average" over all possible formats, something intrinsic about the theme. Let (P_{Mi}) and (P_{Wi}) represent the conditional probabilities of exposure for men and women, respectively, to the them in the i^{th} vehicle and $(R = P_{Mi}/P_{Wi})$ be the (known) ratio of these two probabilities, (at present) assumed constant over all vehicles. Then

---

[10] Actually, it would be a relatively simple procedure to allow different ratios for different vehicles. As presently programmed, the simulation does occasionally call for changes in the ratios when they imply probabilities greater than 1.0.

$P_{Mi} \cdot V_{Mi}$ = Number of men exposed to the theme in the i^{th} vehicle

$P_{Wi} \cdot V_{Wi}$ = Number of women exposed to the theme in the i^{th} vehicle

$A \cdot (V_{Mi} + V_{Wi})$ = Total number exposed to the theme in the i^{th} vehicle and therefore

$$P_{Mi} \cdot V_{Mi} + P_{Wi} \cdot V_{Wi} = A \cdot (V_{Mi} + V_{Wi}) \qquad \text{(VII-1)}$$

with

$$R = P_{Mi} / P_{Wi} . \qquad \text{(VII-2)}$$

For each vehicle, the two equations above are solved for the two unknowns, P_{Mi} and P_{Wi}, and the theme audiences by sex ($P_{Mi} \cdot V_{Mi}$ and $P_{Wi} \cdot V_{Wi}$) can then be calculated. For dimensions which have several levels or categories, there will be correspondingly more ratios to specify, but the procedure and solution to the equations is straightforward.

For any vehicle we have, at this point, a set of consistent message audiences for each level of each dimension for which ratios have been specified. Given these marginal values, it is quite easy to generate message audiences for each of the cells in the table implied by the combination of dimensions. In doing this, we have two choices, however. If we simply take cross-products of these values, we arrive at a set of consistent cell audiences, but audiences which imply no interaction among the dimensions. This, in fact, is the procedure we have used in the simulation.

However, there is another way to generate these cell audiences. We might have initialized the cells with the values of the <u>vehicle</u> audience and performed the smoothing iteration[11] using these initial values. The result of this operation is no longer the production of minimum information or no-interaction cell values, but the generation of cell values which reflect the interaction present in the weighting with which the iteration was initialized.

There are good grounds for considering the initialization of the iteration with the vehicle audience cell values. In the next step, after we have generated the message audience values for the cells, we divide each of these values by the vehicle audience for the corresponding cell in order to get a cell mean, which we interpret as the conditional probability of exposure to the message, given exposure to the vehicle, for a member of the cell described by the total of dimensions for which ratios were specified. In general, it is possible for this procedure to produce message audiences for a cell larger than the cell vehicle audience. This happens occasionally and when it does, we reduce the probability for that cell to .98 and report to the researcher the net effect in decreasing the audience of that message.

---

[11] This was discussed in Chapter II.

However, initializing the cell values with the vehicle audiences would bias the message audiences in the cells in the same way that the vehicle audience is biased. In that way, although the marginals would remain fixed, larger message audiences would occur in cells with larger vehicle audiences and the possibility of getting message exposure probabilities greater than 1.0, would be diminished. It also would imply a leveling of the probabilities of exposure to the message toward some average value, however. At any rate, we have followed the first procedure rather than the second.

There is one last factor which must be added to the model in order to account for different audience levels for different messages even though they occur in the same vehicles carrying the same theme. We call this a "format factor" (F_j is the format factor for the j^{th} message) and it is a positive number which <u>multiplies</u> the average message exposure probabilities derived above, to give the j^{th} message exposure probability. Thus the probabilities derived above for the "average" message format of the theme are probabilities for a message with a format factor equal to 1.0. Formats which increase the likelihood of exposure to the message have format factors greater than 1.0; formats which decrease the likelihood of exposure have format factors less than 1.0. Thus, in the simulation model, in order to specify a scenario of messages for a theme, one must specify the

general average audience for the theme and a set of messages, each consisting of a vehicle number and a format factor which multiplicatively increases or decreases the audience of that message with respect to the average theme audience. In addition, for each vehicle group which the researcher wishes to include, he may specify ratios of conditional exposure probabilities along any number of dimensions of the population.[12] Thus, this is a much simpler body of data for the computer to store and handle, and for the researcher to produce than the most general model.

The model outlined above is convenient for the average researcher who does not have a great deal of data about individual messages. However, for the present simulation, this model implies that we must make at least

---

[12] The product of the largest format factor and the average proportion of the vehicle audience exposed to the theme ($F_{jmax} \cdot A$) gives the largest proportion of the vehicle audience exposed to the theme. This value imposes some constraints on the possible values of the ratios (R). As an extreme example if the maximum message audience is equal to the vehicle audiences, then each conditional probability of exposure must equal 1.0; therefore all ratios must equal 1.0. Obviously, if subgroups for which ratios are specified are quite unequal in size with the ratios implying that the smallest subgroup have the largest conditional probability, then probabilities greater than 1.0 may be generated (and presented questioningly to the researcher) by the simulation if the maximum message audience is a sizeable proportion of the vehicle audience. (There is an option in the simulation to increase and/or decrease vehicle, message, and/or net probabilities according to the time period or previous exposure, or some combination, in order to simulate crisis situations during which the normal probabilities would not be expected to hold. The transformation is almost the same parabolic transformation with ceiling and floor effects described above, p. 255.)

one extra step in preparation of the data for the simulation. In the discussion of the message exposure probabilities generated in the scenario, we have described the generation of conditional probabilities of exposure for males and females for each of the messages in each of the themes. It would be nice if we could use these probabilities directly in the simulation; however, given the model, this is not possible, and we must put these messages and their exposure probabilities into the framework of the simulation model.

Recall that we have for each message in the theme a message exposure probability for both men and women. The model implies that for each message in the theme and vehicle group there is a constant average probability for men and women and that the message audience variation from message to message is due to a format factor which multiplies each of these constant probabilities to give the particular probabilities for the message. Therefore, for each theme, the ratio of our probabilities, the message exposure probability for men and the message exposure probability for women, should be constant over all messages. This is not strictly true for our messages, but, in general, the exposure probability for men is greater than the exposure probability for women and the ratio is nearly constant.

We make the assumption that this ratio is constant enough for us to proceed with the model. In this case, we wish to describe all the messages in a theme and vehicle

group, each with its two exposure probabilities, in terms of three kinds of numbers, i.e., the message exposure probability for men (P_{Mi}), the message exposure probability for women (P_{Wi}), and a format factor for the message (F_j). (To simplify the data processing, we have grouped all the vehicles into one group; thus we may drop the subscript i on the P_{Mi} and P_{Wi}.) Let P_{Mj}, P_{Wj} represent the known (from the linear regression) conditional exposure probabilities for the j^{th} message and P_{Mj}', P_{Wj}' represent the probabilities which will be generated from the proper choice of P_M, P_W, and F_j where P_M, P_W will be the <u>average</u> conditional exposure probabilities for the vehicle group, and F_j the format factor for the j^{th} message. Thus

$$P_{Mj}' = P_M \cdot F_j \qquad \text{(VII-3)}$$

and

$$P_{Wj}' = P_W \cdot F_j \qquad \text{(VII-4)}$$

We want to choose values of P_M, P_W, and the F_j that minimize the difference between P_{Mj} and P_{Mj}', and between P_{Wj} and P_{Wj}'. Thus the problem becomes

$$\text{minimize } \sum_j [(P_{Mj}-P_{Mj}')^2 + (P_{Wj}-P_{Wj}')^2]$$

with respect to P_M, P_W, and F_j. Replacing the P_{Mj}', P_{Wj}' with equations (VII-3) and (VII-4) in the function to be minimized and taking partial derivatives results in the following equations

$$\sum_j (F_j \cdot P_M - P_{Mj}) = 0 \quad \text{(sum over all messages)}$$

$$\sum_j (F_j \cdot P_W - P_{Wj}) = 0 \quad \text{(sum over all messages)}$$

$$F_j = (P_M \cdot P_{Mj} + P_W P_{Wj})/(P_M^2 + P_W^2) \quad \text{(for each message).}$$

We have solved these equations by choosing values of P_M and P_W, calculating the resulting F_j, testing the two sums, adjusting the values of P_M, P_W, etc. This has been done for each of the themes in the scenario, in each case considering all the vehicles as one vehicle group. The outputs are the constant probabilities, i.e., the message exposure probability for men and the message exposure probability for women, and a format factor assigned to each of the messages in the scenario, such that the product of the format factor and the constant probabilities over all the messages comes as close as possible to reproducing the values generated from the regression.

The choice of the average audience (A) for each these is the last remaining problem in transforming the data to fit the computer model. It will be recalled that this average audience is the proportion of the vehicle audience who see the theme on the average, or the audience for a message with a format factor of 1.0. In this case, the message exposure probabilities for men and women will be just the average values generated by the procedure above. If we let V_{Mi}, V_{Wi} represent the number of men

and women in the vehicle audience of the i^{th} vehicle, then this proportion A can be written in the following manner:

$$A = \frac{P_{Mi} \cdot V_{Mi} + P_{Wi} \cdot V_{Wi}}{V_{Mi} + V_{Wi}}$$

(This is just a rearrangement of equation (VII-1) above.)

Now in general, the number of women and the number of men in the vehicle audience are different from vehicle to vehicle; however, it is generally true that the number of women in the audience of any vehicle exceeds the number of men and that the ratio (V_{Wi}/V_{Mi}) lies between the values of 1.0 and 3.0. Therefore, for each theme with its associated average values P_M and P_W, we have calculated the value of A, the proportional average audience, for values of V_{Wi}/V_{Mi} = 1, 2, and 3. These are presented in the table below. We see that the value of A is relatively constant as the ratio varies from 1 to 3. For example, for Theme 1, the value for A varies only from .2000 to .1700. In general, the change in A within a theme for various ratios of V_{Wi}/V_{Mi} is not nearly so large as the change in A for the various themes. In order to simplify matters then, we have used the value of A at V_{Wi}/V_{Mi} = 2.0 for each theme. This is the average value for the proportion of the vehicle audience who see the theme. Thus, for each theme we have produced a ratio of P_M to P_W, a value of A, and a format factor for each message. This has probably

Table VII-5. Values for the Average Proportion of the Vehicle Audience Who are Exposed to the Theme (A) as a Function of the Ratio of the Number of Women (V_W) to the Number of Men (V_M) in the Vehicle Audience and the Average Message Exposure Probabilities for Men (P_M) and Women (P_W)

Theme	Average Message Exposure Probability		Ratio	Average Proportion of the Vehicle Audience Exposed as a Function of V_M/V_W		
	Men (P_M)	Women (P_W)	(P_M/P_W)	$V_M/V_W=1$	$V_M/V_W=2$	$V_M/V_W=3$
1	.2578	.1421	1.810	.2000	.1806	.1710
2	.2808	.1592	1.764	.2200	.1997	.1896
3	.2900	.1502	1.931	.2201	.1968	.1852
4	.3027	.1627	1.860	.2327	.2094	.1977
5	.3202	.1802	1.777	.2502	.2268	.2152
6	.3052	.1152	1.781	.1602	.1452	.1377
11[a]	.1749	.1289	1.355	.1519	.1442	.1404
12	.1850	.1205	1.535	.1528	.1420	.1366
13	.3000	.1682	1.784	.2341	.2121	.2012
14	.3000	.1563	1.919	.2282	.2042	.1922
15	.2753	.1773	1.553	.2262	.2093	.2018
16	.2989	.1695	1.763	.2342	.2126	.2019
17	.3018	.1813	1.665	.2215	.2215	.2114

[a]Computations were not performed for themes 7, 8, 9, and 10, since there themes were not used in the scenario.

caused some distortion in the data; however, this distortion does not seem to be of a magnitude to cause great concern, given the quality of the data before the model was imposed upon it.[13] In the simulation in addition to the ratio value for sex, i.e., the specification of message exposure probabilities for men and women, we have included ratios for age, education, and socioeconomic status (SES). These ratios are derived from data reported by Schramm and White from a 1949 study of a random sample of 746 readers in an Illinois city of a population of 100,00 and for a local evening paper of approximately 65,000 circulation.[14] All the respondents in this study were readers of the newspapers.

Table IX in the study reports the percentage of public-affairs news read, broken down by sex for education,

---

[13] Actually, a more serious distortion may arise from an error in the writing of the scenarios for the simulation which occurred at this point. Inadvertently, the ratio of P_M/P_W for all the themes was set equal to the value of the first theme. This value, as seen from the table, is 1.810 and is in general not too deviant from the values for most of the themes. However, for Themes 11, 12, and 15, this value may cause a significant error in the distribution of exposures between men and women. At any rate, the magnitude of the error attributable to this oversight in writing the scenario is not evident.

[14] The data reported in Wilbur Schramm and David M. White, "Age, Education, and Economic Status as Factors in Newspaper Reading" Journalism Quarterly, (June, 1949), also reprinted in Wilbur Schramm (ed.), Mass Communications (Urbana: University of Illinois Press, 1949). About half the circulation of this evening paper was in the city itself, and the sample of 746 readers was obtained by selecting home addresses throughout the city on a random basis.

Table VII-6. Ratios of Percentages of Public Affairs News Read for Several Population Subgroups.

Dimension	Ratio	Ratio of Percentage for Men to Percentage for Women Within the Dimension Level: Dimension	Ratio (P_M/P_W)
Education		Education	
College / High School	1.176	Grade School	1.696
		High School	1.480
High School / Grade School	1.285	College	1.153
Age		Age	
		10 - 19	.819
20 - 39 / 40 -	.824	20 - 29	1.328
		30 - 39	1.202
		40 - 49	1.133
		50 - 59	1.396
		60 -	1.377
SES		SES	
Highest / Middle	1.303	Highest	1.283
		2	1.243
Middle / Lowest	1.295	3	1.508
		Lowest	1.356

Source: See text above

age, and economic status groups. We note that these percentages of exposure for men and women within each of the classes of education, age, and economic status are summaries over a large number of news stories, each with its own format and themes in terms of our simulation model. Making the assumption that the effects of formating somehow cancel out, or that the net format factor in this case is 1.0, and that the average audience, as a proportion of the vehicle audience, is approximately equal to the figures for our study (or that the ratios hold, even if the average audiences are not the same), we have calculated ratios of conditional exposure probabilities to this news from the table and from the sizes of each of the categories. In Table VII-6 we present these ratios for sex by each of the other three dimensions and for the dimensions of education, age, and economic status.

We see that, according to this study, our estimate of the ratio of sex probabilities is high, that in no category of this study does it reach 1.801, but that it ranges from a low value of .819 in the case of the age group 10-19, to 1.696 for grade school-educated individuals. Omitting the 10-19 age group (which is not included in the simulation population), the average ratio for sex is about 1.30, which is to be compared to the ratios ranging from 1.36 to 1.93 for the themes presented in Table VII-5.

The ratios for age, education, and socioeconomic status as presented in the table were used for the simulation values for each of the themes in the scenario. (This data, incidentally, is corroborated, according to Schramm, by data collected by the University of Minnesota survey on the Minneapolis Star Tribune and by Eugene Liner in an Illinois thesis.) Thus, for better or for worse, we have used what data seemed available and plausible to add additional structure to the media system in the simulation model.

It would be interesting to find out how sensitive the simulation is to changes in the values of these parameters; however, even without any sensitivity testing, we believe that these values realistically reflect the kinds of structure in the real media system.

Anticipating one of the more interesting outcomes of the simulation, let us recall that for the radio vehicles we have specified only the magnitudes of the audiences and the differences in their distribution between males and females. In specifying only these factors and leaving it to the parameter estimation routine to produce the distribution of the audiences for each of these radio vehicles across the population types, we have created exactly similar distributions of the audiences of these vehicles across the other dimensions of the population, i.e., each of the radio vehicles will have the same distribution of men and the same distribution of women across

each of the other dimensions of the population even though the total number of men and the total number of women may differ from vehicle to vehicle. In addition, in choosing values of the ratios that will be constant across all themes, we produce the same relative distribution of message exposure probabilities across the population types for each theme except for the variations introduced by the format factors of the messages. The average probability of exposure for the population to a message in the theme may vary from theme to theme but the relative probabilities will remain much the same from theme to theme across the population types. We shall see the consequences of this fact later in the analysis of the distribution of exposures from the themes.

Triggering and Duplication

When simulating a scenario of a crisis situation, the researcher may often wish to modify the vehicle exposure probabilities, the message exposure probabilities, the format factors, or the products of the three, i.e., the net probabilities of exposure to the message, in order to account for heightened interest in the theme. In the simulation we allow the researcher either to increase or decrease any of these probabilities, either as a function of the person's exposure to some theme or as a function of the time period (e.g., in order to simulate increased attention due to the word-of-mouth spread of information). It is also possible to reverse the

triggering changes or to impose cyclical changes in the probabilities. We have not used any triggering of changes in exposure probabilities in the present simulation, since there was no obvious crisis situation relating to the themes during the period of the simulation; therefore we will not go into the complexities of the triggering mechanisms in great detail. However, it is necessary to understand the notion of exposure classes in order to understand how the duplication statistics are produced in the simulation.

The triggering of probability changes on the basis of previous exposure to some theme or other, requires the retention in the simulation for each member of the population and every time period, of some indication of the person's level of prior exposure to the theme. To maintain the individual's exact number of exposures for every theme and time period would require an impossible amount of storage space in the computer; therefore, we have limited the number of themes which may act as triggering themes to (at most) the first three themes and introduced the notion of exposure classes, rather than exact number of exposures. The researcher may set three levels of exposure, which define the boundaries of four exposure classes (the lower boundary of the first exposure class is taken as zero exposures and the upper boundary of the fourth exposure class as equal to the number of messages in the scenario) and he may request triggering based upon the person's exposure <u>class</u>.

This simplified handling of exposure classes allows the use of only two binary digits (bits) per person to define the exposure classes therefore, for a population of 3,000 people and three themes, the total storage required for one time period is only 504 36-bit words.

The simulation is structured in such a way that in order to produce duplication between themes it is necessary at least to define trigger themes. This arises out of the following problem: in order to make the calculation of the audience duplication between any two themes, one must know for each individual at each time period his cumulative non-exposure probability. Then, one minus this figure is the probability of his being exposed at all to the theme; the product of these probabilities for any N themes is the probability of the individual's being exposed at least once to each of the N themes. The expected number of people exposed at least once to all of the themes considered in the duplication is just the sum of these probabilities for all persons in the cell. In theory this is fine, but in practice it means that one must maintain the cumulative non-exposure probability for every individual, for every time period, for every theme, and possibly by media type, etc. This would require an overwhelming amount of storage and is simply impossible to handle. Therefore, we have settled upon another, cruder method for handling the problem. First of all, we have limited the kind of duplication statistics that

can be obtained to the duplication between any of a set of trigger themes, and any one of the other themes. That is, one may ask for duplication between any trigger theme and any other theme (including another trigger theme) for any time period. Since the cumulative non-exposure probabilities are available for any theme at its running, the model therefore requires additionally only knowledge of the cumulative non-exposure probabilities of one of the triq - themes for the particular time period.

Therefore, we must have some way of getting the cumulative non-exposure probabilities for every trigger theme for each time period, for as many as three trigger themes and any number of time periods. We have allowed the researcher to define three exposure boundaries which define four exposure classes (class 0, 1, 2, or 3) depending upon the cumulative expected number of exposures to the theme through the current time period. The cumulative expected number of exposures is just the sum of all the exposure probabilities. By knowing the class of a person, we can make an estimate of the sum of the number of his exposure probabilities through the current time period. For instance, for a person in the lowest exposure class, we take the average of zero exposures and the number of exposures at the first class boundary as an estimate of his cumulative number of exposures through the time period. (Obviously this is a rough estimate and it is best for the case when the actual

expected number of exposure events for the person lies midway between the class boundaries.) Dividing this estimate of the cumulative number of exposures of the person by the total number of message events through the current time period, gives the estimate of the person's <u>average</u> exposure probability. We estimate the cumulative non-exposure probability from this average. A first estimate assumes that the true exposure probability for each event was equal to the average exposure probability. Each non-exposure probability is one minus the exposure probability and the product of all the non-exposure probabilities is the cumulative non-exposure probability.

We can show that this estimate of the non-exposure probability is the highest possible value: if each of the exposure probabilities were equal to the average, then the product of K individual non-exposure probabilities would be a product of K equal non-negative numbers. If the exposure probabilities were actually distributed about the mean, the product would be a product of K non-negative numbers, some or all of which are not equal to the mean, although the average of these numbers must equal that mean. The first product is the largest possible. If we let P_i $(i = 1, 2, \ldots K)$ represent the K probabilities, then we wish to find the maximum of the product $\prod_{i=1}^{K} P_i$ subject to the constraint that the sum of the probabilities equal a constant $(\sum_{i=1}^{K} P_i = C)$. We form the expression

$$\prod_{i=1}^{K} P_i - \lambda\left(\sum_{i=1}^{K} P_i - C\right)$$

where λ is a Lagrangian multiplier. Setting the partial derivatives equal to zero

$$\frac{\partial}{\partial P_j}\left[\prod_{i=1}^{K} P_i - \lambda\left(\sum_{i=1}^{K} P_i - C\right)\right] = 0 = \prod_{i\neq j}^{K} P_i - \lambda, \; j=1,2,\ldots K \tag{VII-5}$$

$$\frac{\partial}{\partial \lambda}\left[\prod_{i=1}^{K} P_i - \lambda\left(\sum_{i=1}^{K} P_i - C\right)\right] = 0 = \sum_{i=1}^{K} P_i - C \tag{VII-6}$$

we must solve the resulting equations for P_i, i-1,2,...K. We take any two of the K equations (VII-5)

$$\prod_{i\neq j}^{K} P_i = \lambda$$

and

$$\prod_{i\neq 1}^{K} P_i = \lambda \; .$$

Dividing the first by the second

$$\frac{P_1}{P_j} = 1$$

or

$$P_1 = P_j.$$

Therefore the maximum value of the produce occurs when all of the probabilities are equal. The cumulative nonexposure probability thus calculated is an upper bound.

A second calculation assumes that the exposure

probabilities distribute themselves about the average exposure probability which we have just calculated. We generate a symmetric distribution of probabilities about the average exposure probability and calculate another estimate of the cumulative non-exposure probability from this distribution. Since the distribution is widely distributed about the mean, the individual probabilities will be quite unequal. Since the resulting estimate of the cumulative non-exposure probability will be much lower than the previously calculated value and probably lower than the actual value, we take the average of the two calculated values as the best estimate of the non-exposure probability. (We have worked by hand several problems which indicate that this seems to give a good estimate of the non-exposure probability.)

Having calculated the non-exposure probabilities for one of the trigger themes, the probability that a person is exposed at least once to both the trigger theme and any other theme is simply the product of his cumulative exposure probabilities (one minus the cumulative non-exposure probability) for each theme. The sum of these products over all members of the cell is the expected number of people in the cell who have been exposed at least once to both of the two themes.

The definition of the exposure classes serves two purposes then: one purpose is to allow triggering of probability changes based on previous exposure to a given

theme. The second allows the estimation of the non-exposure probabilities in order to calculate duplication between themes. In the first case, the classes would be chosen in such a way as to represent the researcher's feelings about the thresholds of change in attention to one theme because of exposure to another theme. One might imagine the upper boundary of the lowest exposure class being set at about 1.0 expected exposures, if the theme were, for instance, news of the assassination of Kennedy and the researcher anticipated that simply getting this message would be sufficient to significantly heighten exposure probabilities for many other kinds of themes.

However, consider the case for the estimation of the non-exposure probabilities. The first step is essentially an averaging procedure, i.e., the person's mean exposure probability is estimated as the average of the upper and lower bound of his exposure class. The first three exposure classess are bounded on both ends by either 0.0 (for the lowest class) or the values of the boundaries for each class which were chosen by the researcher. However, the upper bound for the highest exposure class can be only the number of messages thus far processed. Therefore, one must use another method of estimating the number of exposures for people who fall in this high class. The estimation procedure is very crude; it amounts simply to assuming that the

average exposure probability was about .5 for each of the exposure events for persons falling into the high class. This assumption, while very rough, probably is not critical for the calculation of the duplication between themes, since the probability of a person's being exposed to the trigger theme, given that he is in the highest class of exposures, is very high regardless of the exact number of exposures.

For the purpose of duplication the boundaries of the lowest exposure classes should form a narrow grid in order to distinguish those people who have relatively little exposure and, therefore, relatively low cumulative exposure probabilities. When one does the mathematics, one finds that the possible range of the cumulative probability of exposure to the theme is very small once the person's cumulative expected number of exposures is at least one. For example, if the person has one expected exposure resulting from one message event, the probability of exposure to the theme is identically 1.0, since his probability of exposure for that particular message must have been 1.0. If the person had one expected exposure in five events so that all the exposure probabilities could have been of the order of .2, the <u>minimum</u> possible

probability of exposure to the theme is .67[15], and even if the person had only one expected exposure in one thousand events, the minimum value for the probability of exposure to the theme is still .63. After a person has (say) ten expected exposures, even for 1,000 message events, his cumulative probability of exposure to the theme is only very, very slightly less than 1.0. Therefore, for the purposes of discriminating between those who are probably exposed to the theme and those not likely exposed to the theme, any number of cumulative expected exposure events greater than one is almost a sure indication that the person is exposed to the theme, i.e., his cumulative probability of exposure to the theme is very high. For this reason, we have set the exposure class boundaries at the low levels, 0.2, 0.6, and 1.2 expected exposures.

---

[15] Let P_i be the probability of exposure to the i^{th} message. Then the non-exposure probability is $1-P_i$ and for five messages, the probability of not being exposed at all is $(1-P_1)\cdot(1-P_2)\cdot(1-P_3)\cdot(1-P_4)\cdot(1-P_5)$. Obviously, if any of the probabilities of exposure equals 1.0, this cumulative non-exposure probability is equal to zero; it has a maximum value when all the P_i are equally small. Now the number of expected exposures (by assumption, 1.0) is just the sum of the five P_i; therefore the values of the P_i which maximize the cumulative non-exposure probability are 0.2 (i.e. five equally small values which sum to 1.0) and the maximum cumulative non-exposure probability is 0.32768 (0.8^5). Thus the minimum cumulative probability of being exposed at least once is 0.67232 (1.0 - 0.32768).

CHAPTER VIII

THE SIMULATED EXPOSURES TO THE CINCINNATI MASS MEDIA SYSTEM

In the previous chapters, we have specified the model to the Cincinnati mass media system, creating the vehicles and their audiences, and we have recorded the actual messages in the system over the six months with their message exposure probabilities. At this point, we are ready to run the simulation. In the first part of this chapter we examine the consistency with which the model has synthesized the input data, by use of exposure statistics from three simple trial scenarios. We will find that the model does reproduce the input data and produce plausible exposures to these three scenarios. Next, we will look at the exposure results for the twelve themes in the scenario, examining the growth and distribution of exposure, its duplication across themes, and the relative importance of the various media vehicles. Finally, we will attempt to correlate exposures to the twelve themes with indices of actual change from the NORC panel. We close with several comments about the relative importance of the different kinds of input data, some suggestions toward increasing

the usefulness of the model, and an admonition or two for designers of simulation models.

Validation of Several Parts of the Simulation Model

If the simulation model--our consistency machine--is to be of any use in predicting exposures of populations to messages in the communication vehicles, then it should at a minimum be consistent with the input data which specify the parameters of the model. The simulation should, in addition to synthesizing and integrating the input data, also reproduce or describe accurately the input data. This is, in Hanna's terms, the descriptive validity of the model.[1] In this section, we will look at how well the parameter estimation routine, cumulation routine, the assignment of probabilities to cells, and the duplication routine reproduce that known data which was input in order to specify the model to a given population and mass media system. We begin with the parameter estimation iteration in the first link of the simulation.

The Mosteller Parameter Estimation Iteration

Does the iteration process for integrating partial data succeed in reproducing a population and vehicle audiences which duplicate the input data? The answer to this

---

[1]Hanna, Information - Theoretic Techniques

question is, yes, the iteration process does produce consistent population and vehicle audiences within the margin of inaccuracy which one finds in the input data. Let us be more explicit about the inaccuracies in the input data. The population tables which were used to create the simulation population consisted of all the three dimensional tables possible from the five population dimensions. There are ten of these three-dimensional tables and they are not totally consistent with each other because of non-response or not ascertainable responses in the survey data. For example, the total population in these subtables ranges from 1097 respondents to 1112 respondents. Thus, we have a variation of at most 15 in the sum of the frequencies in the subtables. Since these tables of about 1100 respondents are used to generate a population of size 2000, meaning that each of the values in the original table must be multiplied by a factor of about 1.8, we see that there are significant inconsistencies possible in the cells of these subtables. The iteration routine will produce a population which has an average error which may be anywhere from half as large to just as large as the discrepancies in the input tables. These errors are then multiplied by the factor of 1.8 in normalizing the population to the value of 2000, and then perhaps some slight additional error is made when the nonintegral values are rounded to produce integral numbers of respondents in each

Table VIII-1. A Comparison of One Subtable of Population Values from the NORC Survey Data with the Synthesized Data Produced by the Parameter Estimation Routine

	Males, of Education				Females, of Education			
Interest	College	High School	Grade School	Total	College	High School	Grade School	Total
NORC Survey Data (Normalized to 2000)								
Low	16.19	64.75	86.33	167.27	7.19	133.09	179.85	320.13
Low-middle	75.54	102.51	98.92	276.97	16.19	154.67	111.51	282.37
High-middle	75.54	73.74	64.75	214.03	43.16	122.30	57.55	223.01
High	75.54	134.89	70.14	280.57	55.75	124.10	55.75	235.60
TOTAL	242.81	375.89	320.14	938.84	122.29	534.16	404.66	1061.11
Simulation Produced Data								
Low	13	70	73	156	7	145	183	335
Low-middle	64	117	100	281	28	142	112	282
High-middle	74	73	61	214	40	118	56	214
High	85	1[illegible]2	62	269	49	139	61	249
TOTAL	236	388	296	920	124	544	412	1080

cell of the population. Therefore, in evaluating the success of the iteration routine, we must remember that the input values themselves are not consistent, but may have subtable cells which are inconsistent by as much as ten or fifteen respondents which under the normalization must be multiplied by a factor of 1.8. Therefore, we expect imperfect agreement but some sort of average values for the population values produced by the iteration routine, the normalizing to the simulation population, and the rounding to the cell values. In Table VIII-1 we compare one of the input tables from the NORC survey data with the population values produced by the simulation. In order to facilitate the comparison, the population values from the survey data are normalized to a total population of 2000.

In comparing the two tables, we see that for some of the large cells the simulation values are larger than the survey data by 15 respondents or about 15 percent. Note, for instance, the differences in the two values for the males with high school education and low-middle interest. In the light of our discussion above, we must not make the mistake of considering this difference between the simulation and the input values as due totally to an error in the simulation population values, since the input tables are themselves inconsistent and the iteration produces a population which attempts to reconcile so far as

Table VIII-2. A Comparison of One Subtable of *Enquirer* Audiences from the NORC Survey Data with the Synthesized Data Produced by the Parameter Estimation Routine

	Males, of Education				Females, of Education			
Interest	College	High School	Grade School	Total	College	High School	Grade School	Total
NORC Survey Data (normalized for a population of 2000)								
Low	14.72	21.26	21.26	57.24	3.27	39.24	35.97	78.48
Low-middle	37.61	49.05	42.51	129.17	9.81	52.32	35.97	98.10
High-middle	39.24	24.53	13.08	76.85	24.53	40.88	14.72	80.13
High	63.77	60.50	26.16	150.43	26.16	67.04	21.26	114.46
TOTAL	155.34	155.34	103.01	413.69	63.77	199.48	107.92	371.17
Simulation Produced Data								
Low	11.37	26.11	28.36	65.84	4.77	35.07	33.29	73.13
Low-middle	36.02	47.16	35.29	118.47	16.44	56.44	39.67	112.55
High-middle	37.95	27.43	14.91	80.29	18.88	41.65	16.94	77.47
High	62.26	51.83	28.13	142.22	21.29	64.84	25.39	111.52
TOTAL	147.60	152.53	106.69	406.82	61.38	198.00	115.29	374.67

possible the inconsistencies in the input data. At any rate, a comparison of the two tables seems to show that the iteration, normalizing, and rounding process fairly well duplicates the distribution of the population across the cells as given by the input data.

In Table VIII-2 we make a similar comparison between the audiences in the population subgroups of the morning Enquirer as generated by the NORC survey data and the same audiences after the iteration technique has synthesized the ten survey subtables.

The table of audiences indicates about the same order of error as in the previous table for the population data. Again, we note that the differences between the values generated by the simulation and the values input from the survey data are not entirely error, since the simulation is attempting to reconcile inconsistencies in the input data. The parameter estimation technique does seem to synthesize the input tables correctly. Of course, what it cannot do is produce any unknown higher order interactions.

Trial Scenarios to Validate the Cumulation and Probability Assignment Routines

To test whether the simulation was correctly synthesizing the input data (in the sense of being able to reproduce it, at a minimum) and to judge the plausibility

of its output, we ran several simplified trial scenarios before proceeding to the actual content of the Cincinnati mass media system. In order to plot the penetration of the media vehicles in a single weekday, the first scenario divides the day into thirteen time periods and uses a single story in each of the vehicles throughout the day. We could describe this scenario as a one-day average news story. A second trial scenario represents a three weekday average story, dividing each day into six time periods (for a total of eighteen time periods over the three days) with a single message in each vehicle each day. The third trial scenario represents a more important three-day news story. This scenario differs from the average three-day story in that each newspaper now carries three messages relating to the story at each appearance of the newspaper over the three days.

The trial scenarios were designed not only to test the effect of news events of differing magnitude and duration, but also to display the effect of the message exposure probabilities. Recall that the probability distributions over the population produced by the first half of the simulation, attempt to reproduce the vehicle audiences: it is in the second half of the simulation with the actual messages, their format factors, the ratios of exposure probabilities by sex, age, etc., that we can calculate message exposure probabilities which are actually conditional probabilities of exposure to a message in a

vehicle, given exposure to the vehicle. Thus, by running each of the trial scenarios with message exposure probabilities equal to 1.0, we produce the vehicle audiences as output; if we run the trial scenarios with more realistic (smaller) message exposure probabilities, we produce audiences of a series of messages in these vehicles. For each of the trial scenarios above, we have used both sets of message exposure probabilities,[2] producing in effect six scenarios, three of which report the vehicle audiences of each of the news events and three of which report plausible exposures to messages of a news event as carried in the vehicles.

To summarize, each of the following hypothetical scenarios is run both for the vehicle audiences and for the actual news event audience:

1. One-day average story. One message in each vehicle during the day, and thirteen time periods.
2. Three-day average story. One message in each vehicle for three consecutive weekdays, six time periods per day.
3. Three-day larger story. One message in each radio vehicle and three messages in each newspaper for three consecutive days, six time periods per day.

---

[2]The realistic message exposure probabilities and ratios are the average values calculated in Chapter VII.

Table VIII-3. Occurrences of Vehicles and Numbers of Messages by Time Period for One-day and Three-day Scenarios

One-day Scenarios			Three-day Scenarios			
Period	Vehicle Occurrence	Total Number of Messages of the Time Period	Period	Vehicle Occurrence	Total Number of Messages in the Time Period: Average Story	Total Number of Messages in the Time Period: Big Story
1	7 AM radio	1				
2	8 AM radio, Enquirer	2	1	7,8 AM radio, Enquirer	3	5
3	9,10 AM radio	2	2	9,10 AM radio	2	2
4	11 AM radio	1	3	11,12AM,1 PM radio	3	3
5	12 AM radio	1	4	3,4,5 PM radio, Time-Star, Post	5	9
6	1 PM radio	1	5	6,7,8 PM radio	3	3
7	3,4,5 PM radio, Time-Star, Post	5	6	9,11,12 PM radio	3	3
8	6 PM radio	1		Total number of message per day	19	25
9	7 PM radio	1				
10	8 PM radio	1				
11	9 PM radio	1				
12	11 PM radio	1				
13	12 PM radio	1				
	Total of messages in the scenario	19		Total of messages in the scenario	57	75

Table VIII-4. Format Factors and Ratios for Realistic Message Exposure Probabilities

Vehicle	Format Factors	Message Exposure Probability Ratios		
		Sex	Education	Interest
Morning Enquirer	.172	$\frac{P(\text{Male})}{P(\text{Female})} = 1.616$	$\frac{P(\text{College})}{P(\text{High School})} = 1.3$	$\frac{P(\text{Low})}{P(\text{Low-Mid})} = .75$
Times-Star	.169			
Post	.169		$\frac{P(\text{High School})}{P(\text{Grade School})} = 1.3$	$\frac{P(\text{Low-Mid})}{P(\text{High-Mid})} = .80$
A 1 Radio News-broadcasts	.5			$\frac{P(\text{High-Mid})}{P(\text{High})} = .71$
(PORTN - 1.0)				

Table VIII-3 and VIII-4 show the occurence of vehicles and messages by time period and the realistic format factors and ratios.

Before we examine the growth of vehicle and message exposures in the trial scenarios, we might first ask if the simulation is correctly reproducing various aspects of the input data such as the average audiences and cumulations of the vehicles, and the distribution of the audiences over the population types.

Both of these questions can be answered affirmatively by comparison of the output of the trial scenario with the input data. In the trial scenarios the Enquirer is defined as the only vehicle in media type 1 (morning newspapers). Thus, by referring to the output tables describing exposure by media type, we can observe the cumulative growth of the vehicle audience of the Enquirer (in the scenarios with message exposure probabilities equal to 1.0). Table VIII-5 shows the relevant data. The input audience for the Enquirer was about 39.20 percent of the population. The parameter estimation routine compromised the slight inconsistencies in the input data and distributed an audience of 39.24 percent across the population cells. The cumulation routine was given this value and input data describing the two-period cumulative audience of the Enquirer as 45.51 percent of the population. It generated 2000 discrete probabilities which produce a

final output expected audience of 39.35 percent and a final output expected two-period cumulative audience of 45.51 percent of the population.

Table VIII-5. Comparison of Input Data and Simulation Output Data for the *Enquirer*.

Statistic	Input	Parameter Estimation Routine	Output from Cumulation Routine	Output from Trial Run (After Probability Assignment)
Average Audience	39.20%[a]	39.24%	39.35%	39.35%
Two-Period Cumulation	45.51%		45.51%	45.51%

[a]Data are audiences as percentage of the population.

From the *Enquirer* output data we can also judge the success of the simulation in reproducing the cell-by-cell average audiences produced by the parameter estimation process. After the cumulation routine generates the 2000 discrete probabilities, these are then assigned to the cells of the population, one probability for each member of the population. This assignment is random except for the constraint that the assignment attempts to produce an expected cell audience equal to that required from the parameter estimation routine. Table VIII-6 shows the actual output from a trial scenario for the *Enquirier*.

Table VIII-6. Expected Vehicles Audiences of the *Enquirer* (Output of Trial Scenario One with Message Probabilities Equal to 1.0)

	Males				Females			
	Education				Education			
Interest	College	High School	Grade School	Total	College	High School	Grade School	Total
Low	9.50	25.66	27.96	63.12	4.99	35.92	34.04	74.95
Low-middle	33.92	47.52	36.80	118.24	15.52	56.62	39.94	112.08
High-middle	38.74	27.74	16.38	82.86	19.16	41.52	17.18	77.86
High	61.80	52.64	28.42	142.86	21.24	66.64	24.80	112.68
TOTAL	143.96	153.56	109.56	407.08	60.91	200.70	115.96	377.57

These data should be compared with the results of the parameter estimation routine in Table VIII-2, showing the same audience as originally input and as produced by the parameter estimation. The two tables of audiences are very close, the largest absolute discrepancy of 2.10 persons occurring for the cell of males of low-middle interest and college education. Also, comparing the marginals of the tables, we see that as we aggregate or collapse across cells, the discrepancies become smaller as they tend to cancel each other.[3] On the basis of this comparison we conclude that the process of generating and assigning discrete probabilities does faithfully reproduce the data presented to it.

The Vehicle Audience Duplication

The final important segment of the simulation which we have not yet examined for validity is the method of attempting to reproduce empirical duplication figures by the assignment of probabilities to individuals within the cells.

[3]Actually the dimensions of the population define 144 population types or cells, and the random assignment process attempts to reproduce the parameter estimation resulting audiences for each of these cells. Since the average size of each of these 144 cells is one-sixth (144/24) the size of the 24 cells presented above, the sample of probabilities drawn for any cell is only one-sixth as large as for the 24 cells. Therefore we would expect the proportional deviation of the expected audience from the required values to be greater for the 144 cells than for the 24 cells. However, this matters little, since we will always collapse the 144 cells to some much smaller number for output and analysis.

This assignment is another feature which adds structure to the simulation: it defines the audience duplicated between two vehicles and thus the reach and frequency of exposure of various combinations of media vehicles. Unfortunately, the data in the trial scenarios do not allow us to separate out the effects of the within cell non-random assignment of probabilities in structuring the duplication of the population; the scenarios would have to be more simple than those which we have run in order to test the effect of the nonrandom within cell assignment. However, there is another structure which we suspect is much more important in determining the duplication between any set of vehicles, namely, the assignment of average audiences to the cells of the simulation. The choice of the dimensions of the population was based on the effect of these dimensions upon the audience of the vehicles, i.e. the dimensions were chosen to define the greatest differences between population types in their exposure to the vehicles of the mass media simulation. In effect these dimensions are the dimensions which most account for the variability in exposure of individuals to the mass media; therefore we might suspect also that these dimensions account for much of the duplication between any pair of vehicles. For example, if there are two vehicles, vehicle A and vehicle B, each of which reaches one-half of the population in its average audience

and if the distribution of the audiences of these vehicles were completely random across the population, then we would expect that on the average 25 percent of the population would be exposed to both the vehicles and that the combined audience of the two vehicles or the coverage of the population by the two vehicles would be 75 percent of the population. This is the expected audience based on an absolutely random audience distribution throughout the population. Imagine, however, a different situation. Suppose those people who are subscribers or quite regular readers to vehicle A are also the people who are subscribers and quite regular readers of vehicle B; likewise the infrequent readers of one vehicle are the infrequent readers of another vehicle. In this case we expect a greater duplication of the audiences of vehicle A and vehicle B than from the random model. Some number greater than 25 percent of the population would be in the audience of both vehicle A and vehicle B and correspondingly the coverage of the two vehicles would be less than 75 percent of the population.

This kind of situation is largely accounted for by the distribution of the audiences over the cells which type the population: since the population dimensions serve to explain much of the readership of a vehicle, two vehicles which empirically have a higher than random duplication will tend to have large audiences or high

probabilities of exposure in the same cells or population types. Conversely, two vehicles which have lower than average random duplication will probably have their high audiences in different population types. Therefore, we expect that a major part of the nonrandom duplication found between any two vehicles will be explained (and produced) simply by the relative distributions of the audiences of the two vehicles over the population types.

We have tested this expectation for an extreme case with two of the vehicles in the trial scenarios, the Cincinnati evening newspapers. Each of these newspapers has a very high average audience and on the basis of the random audience distribution model, we would expect to find a very high duplication of audience between the two. We can even argue that in the simulation and/or empirically we might even expect larger than random duplication since the readers of one evening newspaper might be much like the readers of another evening newspaper in their distribution by sex, age, interest, socioeconomic status, and education. Of course we find, on the contrary, that the empirical duplication is much <u>lower</u> than the duplication one would expect simply by chance; the readers of one evening newspaper do not generally read the other evening newspaper. Thus, if for these two vehicles for which there is a large difference between the empirical and chance duplication, we find that the distribution of

the two audiences across the cells accounts for most of the difference between the empirical and the random duplication, then a fortiori we can assume for other vehicles that, even before the assignment process, the cell audiences have probably accounted for most of the nonrandom duplication found between the two vehicles.

The audiences of the Times-Star and Post are 50.2 percent and 47.6 percent of the population respectively. If these two audiences were randomly distributed across the population, we would expect a chance duplication based simply on these audience sizes equal to 23.8 percent of the population, or that 23.8 percent of the population is exposed on the average to both the evening newspapers. The empirical duplication, however, is 5.2 percent of the population; as might be expected, the actual duplication between the two evening newspapers is much lower than an estimate based on a purely random distribution of the audiences across the population. What proportion of this nonrandom duplication is accounted for by the two audience distributions over the five dimensions? This expected duplication, based on the individual cell audiences of the two vehicles, is just the sum, over the 144 cells or subgroups defined by the five dimensions, of the duplication within each cell. This expected audience of both evening newspapers divided by the simulation population, gives the proportion of the population in the

audience of both the vehicles. The trial scenarios do not produce, for each of the evening newspapers, the cell audience actually achieved by the random assignment of the discrete probabilities produced by the cumulation routine. We know, however, that these probabilities produce average audiences which are quite close to the value generated by the parameter estimation routine for the 144 cells; therefore we have used these values, which are available, in the calculation of the duplication between the two vehicles based simply upon the average audiences in the cells. The result of the calculation is a duplication of 7.28 percent of the population for the two evening newspapers. Thus the simulation distribution of the audiences of the two vehicles over the population types, based, of course, on the input data, has produced a duplication between the two vehicles which comes quite close to the empirical duplication, 5.2 percent, the difference being only 2.1 percent of the population. Therefore, we conclude that even if the within cell assignment of probabilities to adjust for the nonrandom duplication not accounted for by the cell audiences is a total failure, the distribution of the audiences across the population types chosen for the simulation will most likely produce duplication values which are quite close to the empirical findings.

The Internal Validity of the Model

Charles Herman, in a paper on validation problems in games and simulations defines the concept of the internal validity of a model:

> The unexplained variance between intended replications would provide a measure of reliability or what Campbell. . . calls "Internal Validity." When the structured simulation properties are held constant, the smaller the between run variance, the greater the internal validity is assumed to be.[4]

This concept of validity does not apply in an absolutely straightforward way to the present simulation since the output of the simulation is in the form of expected values of the random variables. In order to judge the internal validity of the model, we must know the range of possible outcomes for different runs given identical parameter values. An estimate of the variance of the random variables, in addition to the expected values, would be required. Because of space and time limitation we have not programmed the simulation to produce these variances, although this is quite easy to do mathematically. However, we can report on the variability in the expected values due to the random probability assignment.

The expected values of the random variables which are

[4]Charles F. Herman, "Validation Problems in Games and Simulations with Special References to Models of International Politics," Behavioral Science, Volume 12, 1967, pp. 216-231.

output by the simulation are really expected values conditioned upon the random assignment of probabilities to the cells and the pairing of these probabilities to account for the duplication. In theory we could generate a probability distribution for all possible assignments of probabilities given the assignment process and combine that with the message exposure probabilities to generate unconditional expected values at the output stage of the simulation; however, this theoretical possibility is mathematically quite complex and we have not attempted to do this. Therefore, we are left with the internal validity problem: what is the range of the conditional expected values of the output variables given other runs of the simulation and with other assignments of the discrete probabilities? For the total population such statistics as cumulation, the average number of exposure events, and the proportion of persons exposed at a given time period are constant for different random assignments of the probabilities, since the distribution of the discrete probabilities generated by the cumulation routine is completely determined. Therefore, any statistics for a single vehicle and the entire population at every time period must remain constant regardless of the assignment process. However, the same statistics need not remain constant for a vehicle within a given cell since the random assignment

process will in general produce different discrete probabilities in each cell of the simulation.

A comparison of the expected audience of the Enquirer as output from the parameter estimation routine with that produced by the probability assignment (Tables VIII-2 and VIII-6) gives some indication of the variability which we might expect at the level of the subgroups which we typically output from the simulation. The random assignment process attempts to produce in each cell an expected audience which was arrived at by the parameter estimation routine. The amount by which the particular assignment which we have observed differs from these expected values gives us a measure of the variability which we might expect from the random assignment process, i.e. if we look at the differences between the values required by the estimation process and the values actually achieved by the random assignment, unless this random assignment is a very peculiar one (which would be quite unusual over the twenty-four subgroups in Tables VIII-2 and VIII-6), then these differences indicate the amount of fluctuation we can expect in the cell by cell average audiences due to the random assignment process. The tables show that the amount of this error is at most 2.1 persons exposed in one cell and, in general, that the error is very small. Therefore we may conclude that the variability produced by the random assignment process does not affect the

expected audiences in the output subtables by more than a few percent.

How does the expected cumulative audience of a vehicle in the individual cells differ from random assignment to random assignment? We don't know the answer to this, but since the average audience seems to exhibit little variability from random assignment to random assignment we speculate that the cumulative audiences will also exhibit small variability from random assignment to random assignment, although admittedly that variability may be magnified as the number of time periods in the cumulation grows larger and larger.

The foregoing discussion has been in terms of the possible audiences variations in various population types for a single vehicle. We saw that the average and cumulative audiences of the vehicle over the entire population would not change given a different random assignment of probabilities but that there might be some slight differences in the expected audiences for a given cell of the population given a different assignment of probabilities and there may be some variation, although hopefully still a small variation, in the cumulative audiences for a given cell or subgroup. Also these variabilities due to sampling will be smaller the larger the population of

the cell or the smaller the dimensionality of the subtable containing the output.

What can we say about the variability in the average and cumulative audiences of a theme of messages appearing in several different vehicles? Obviously the exposure of an audience to a theme of messages involving several vehicles depends in part upon the duplication between the vehicles, which is to some extent manipulated by the process of assigning probabilities across vehicles within the cells. The particular probabilities which are randomly assigned to the cells may affect the ability of the simulation to manipulate the duplication to conform to the empirical duplication. This variability in duplication from assignment to assignment is something which we have no measure of; however, we note from the discussion of duplication above that by far the largest amount of duplication across vehicles is accounted for by the average audiences of these vehicles within each cell of the simulation. These mean audiences are fixed by the parameter estimation process and do not fluctuate substantially given random assignment of probabilities as we have seen from our comparison of Tables VIII-2 and VIII-6 above. Thus we conclude that the largest proportion of the duplication will remain relatively constant from random assignment to random assignment, but that some proportion of the already small proportion of the duplication

which is manipulated by the assignment process will change from random assignment to random assignment. Thus, although the exposure of the audience within a cell or over the entire population to a theme of messages involving several vehicles will fluctuate from random assignment to random assignment, we offer a rough estimate that this variability in the "expected" audience will not be greater than a few percent.

Some Final Thought About Internal Validity

Although we have introduced the notion of internal validity and, in keeping with the spirit of the concept, attempted to make some assessments of the between run variability in the simulation (noting, however, that the most important variance was not calculated in programing the model), we admit to some second thoughts about the implications of the concept. The concern, stimulated by Hanna's ideas about validating models, is that a model may have <u>too little</u> between run variability.

What is the relationship between the observed data and reality? We consider that the observed data are produced by a combination of the factors which the model attempts to account for, other (exogenous) factors which the model does not attempt to account for, and measurement error. If the real world could be "rerun" (as the simulation can), controlling for the simulation factors,

but allowing all possible variation in the unaccounted for factors and measurement error, then we could produce a probability distribution for the possible observed outcomes. The simulation is valid insofar as its probability distribution of simulated outcomes matches this hypothetical probability distribution of real world outcomes.

Referring to the exogenous factors and measurement errors as the loci of ignorance of the model, Hanna shows that "...the most informative model is one whose predictions have exactly the degree of precision permitted by the loci of ignorance associated with the referent situation. Thus, one can lose information due to over-specific as well as the under-specific predictions."[5]

The Results of the Trial Scenarios

We turn now to the graphs depicting the growth of exposure to the trial scenarios among the various population subgroups. The graphs in Figures VIII-1 through VIII-3 show the growth of the cumulative percentages of males and females exposed at least once to the scenario. The upper points represent exposure to vehicles, the lower points exposure to messages in these vehicles. For each of the scenarios, the percentages exposed at least once to the vehicles very quickly approach 100 percent. Thus

---

[5] Joseph F. Hanna, Information-Theoretic Techniques for Evaluating Simulation Models, Michigan State University, 1969, Mimeo, pp. 11-12.

for the first trial scenario, the vehicle audience rises to greater than 95 percent by the end of the thirteenth time period, which is the last radio message in the day. There are two points where the graph rises quite rapidly--in the second time period when the mroning _Enquirer_ is available and in the seventh time period when the _Post_ and the _Times-Star_, the two evening papers, become available. It is quite likely that these increases should not be so abrupt, that, in fact, the exposure takes place over several hours after the newspapers hit the newsstands; however, we have no way of taking account of this in the simulation. At any rate, by the end of the day at least 95 percent of both men and women are exposed to a vehicle at least once.

The graph of the cumulative percent exposed to a message is similar in shape to that of the curve for the cumulative number exposed to the vehicle, but the cumulative percent exposed at each time period is significantly less. Thus at the end of 13 time periods, for the first trial scenario the cumulative percentage of males and females exposed at least once to the message is about 66 and 54 percent, respectively. One effect of the more realistic message exposure probabilities is to make a larger difference between the proportion of males and the proportion of females exposed to the message.

Because the cumulative percentage of people exposed

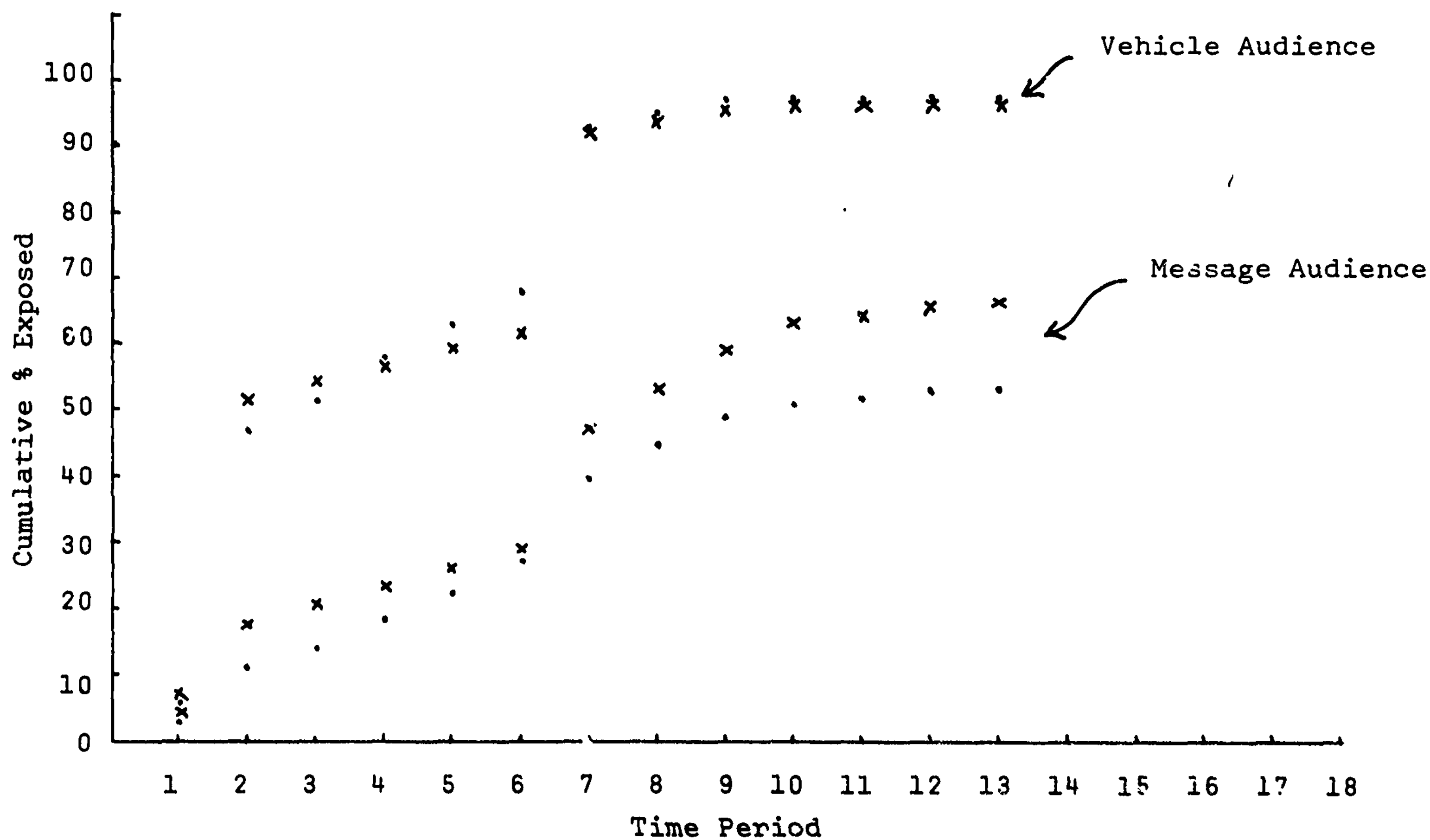


Figure VIII-1. The Cumulative Percentage of Males (x) and Females (.) Exposed at Least Once to the First Trial Scenario

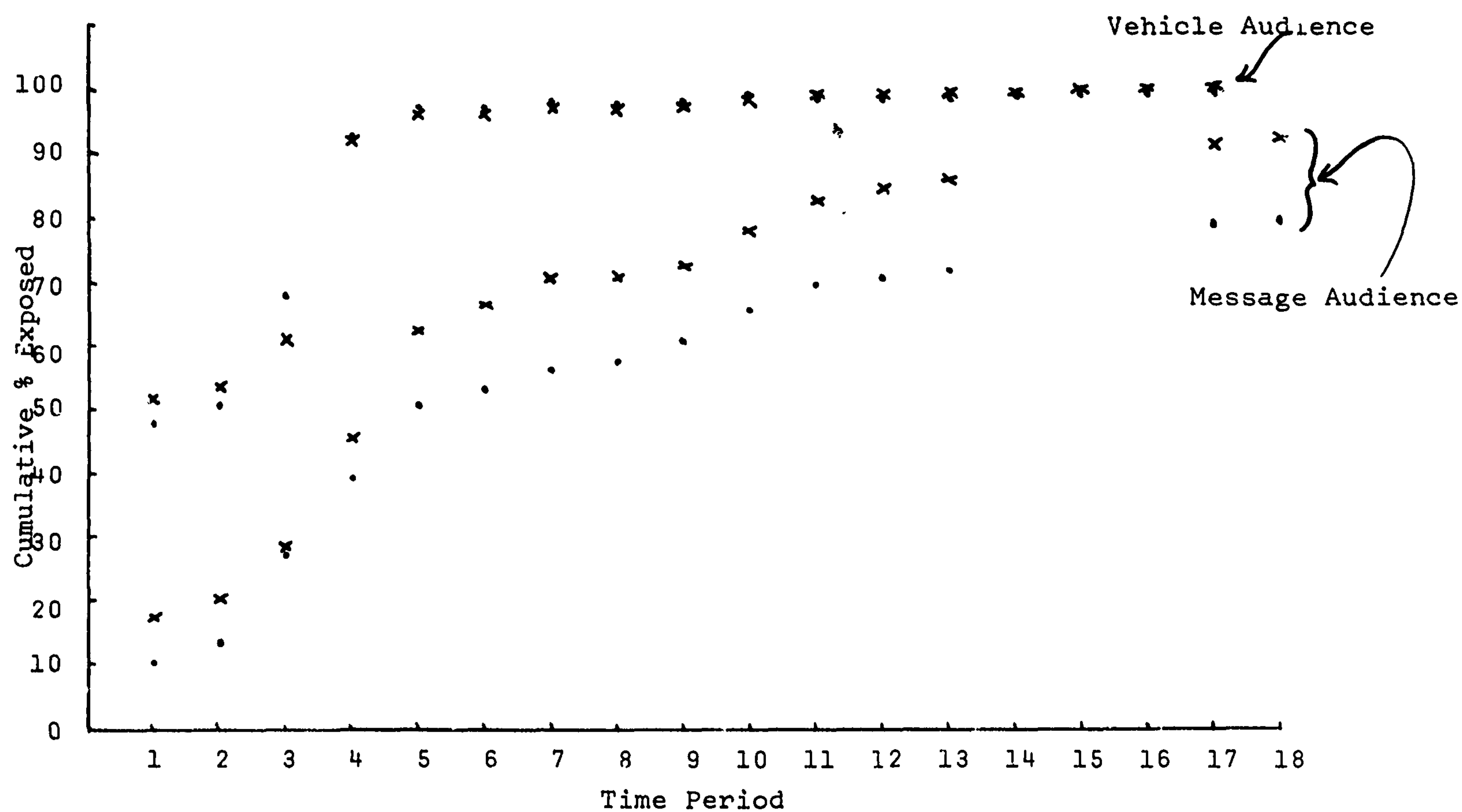


Figure VIII-2. The Cumulative Percentage of Males (x) and Females (.) Exposed at Least Once to the Second Trial Scenario

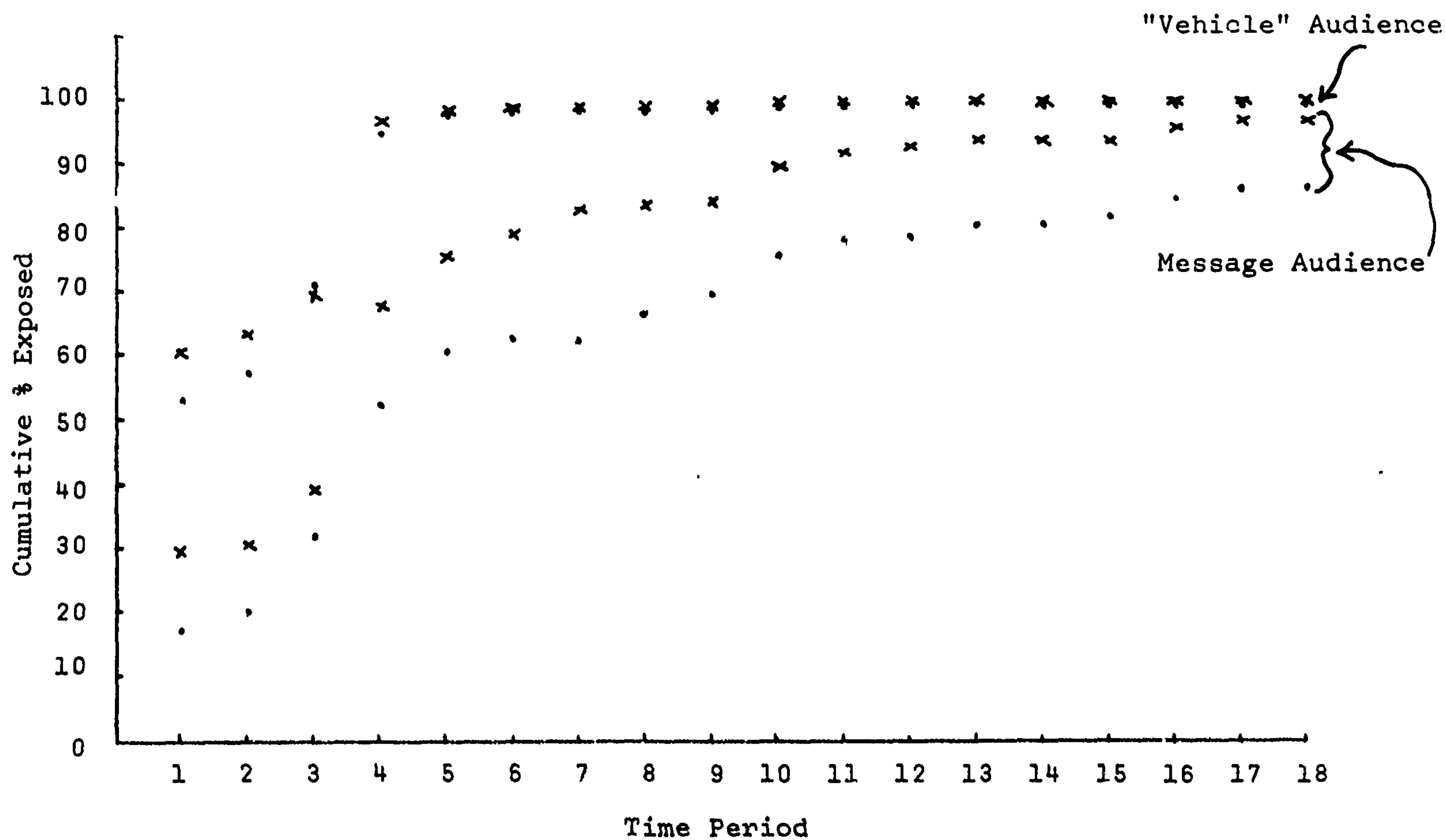


Figure VIII-3. The Cumulative Percentage of Males (x) and Females (') Exposed at Least Once to the Third Trial Scenario

grows so quickly toward 100 percent, we have plotted a second set of curves (Figures VIII-4 through VIII-12) representing the growth of the average cumulative number of exposures for the subgroups defined by sex, education, and socio-economic status. This is a much more important measure of exposure since it seems likely that a certain minimum number of exposures is necessary in order for a person to retain or be effected by the content of the average message. We observe an interesting effect in the graphs by sex. For the vehicle audiences, women are consistently more exposed than men; however, when the message exposure probabilities are used, the men are the more exposed. Of course, none of the differences in exposure between subgroups is very large (a maximum of about 1.5 exposures by the eighteenth time period) for these short scenarios.

Table VIII-7 summarizes the results of the three trial scenarios after the final time period for each theme. This table presents data on the average of the cumulative number of exposure events for various subgroups in the population. The first of each pair of rows describes the vehicle audiences and the second the message audiences. Since the second theme is equivalent to the first theme run three times, we will look particularly at the first and the third themes for our analysis.

First, the message audiences are about one-quarter

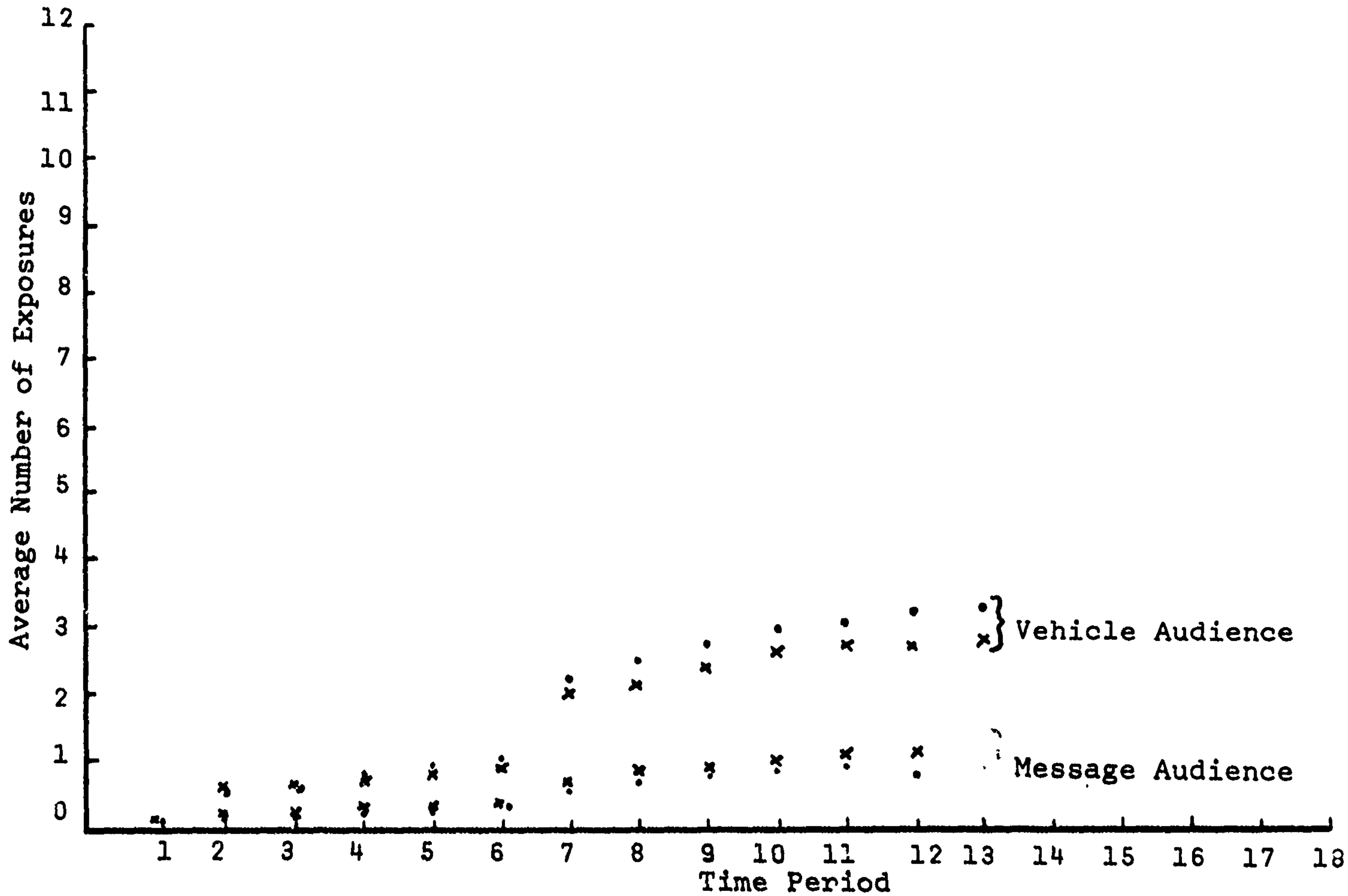


Figure VIII-4. The Average Cumulative Number of Exposures of Males (x) and Females (•) to the First Trial Scenario

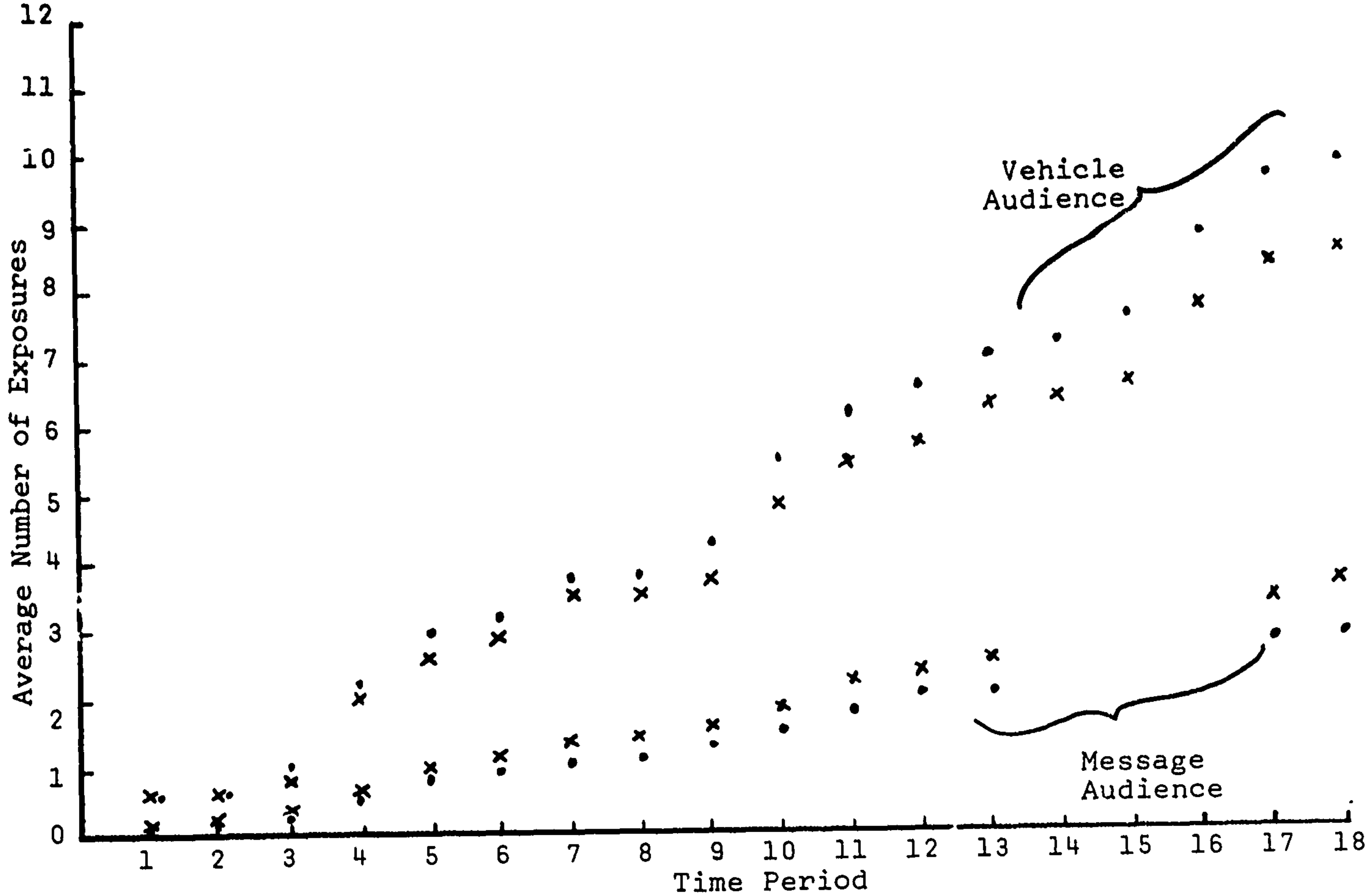


Figure VIII-5. The Average Cumulative Number of Exposures of Males (x) and Females (•) to the Second Trial Scenario

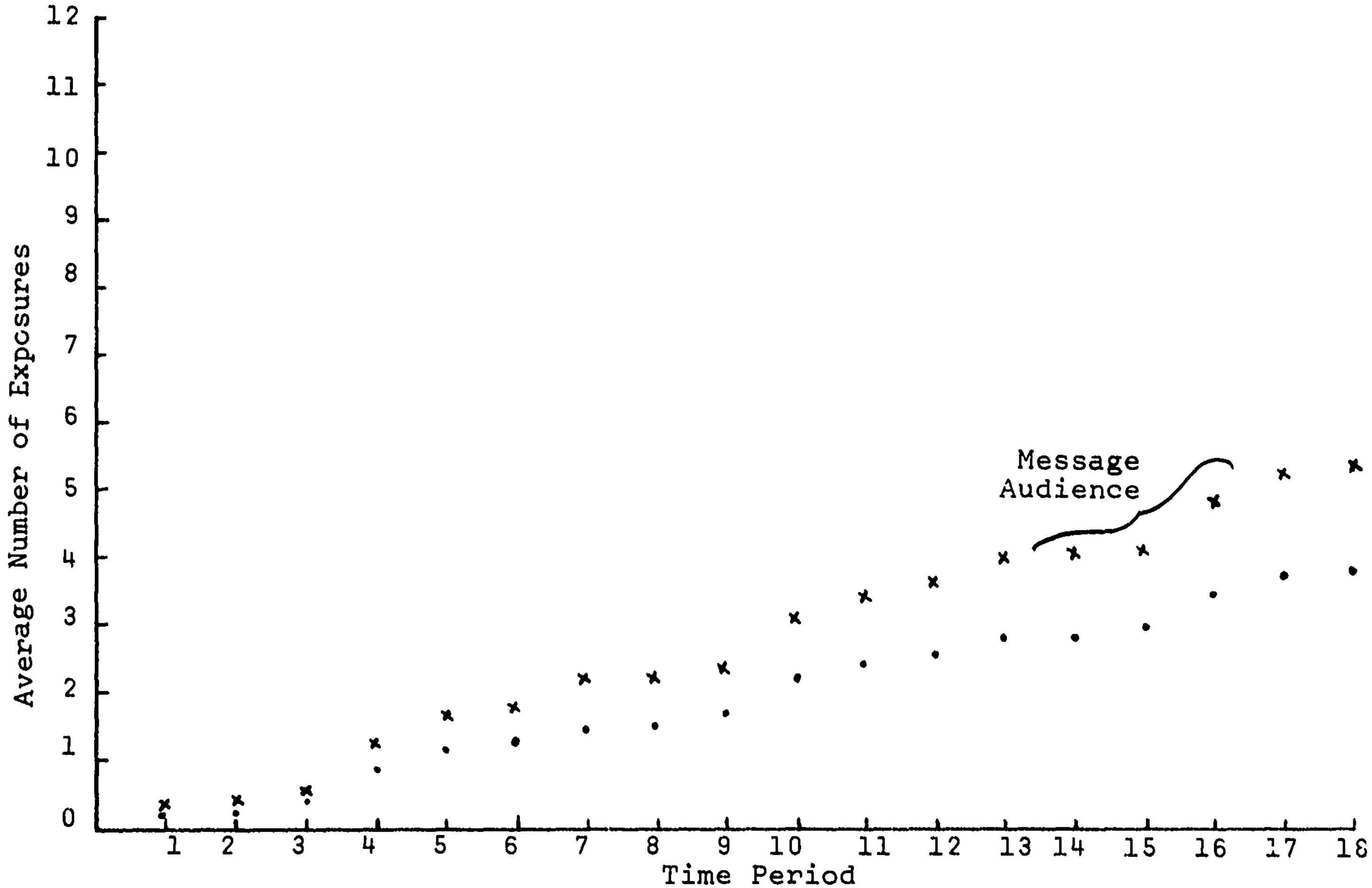


Figure VIII-6. The Average Cumulative Number of Exposures of Males (x) and Females (•) to the Third Trial Scenario

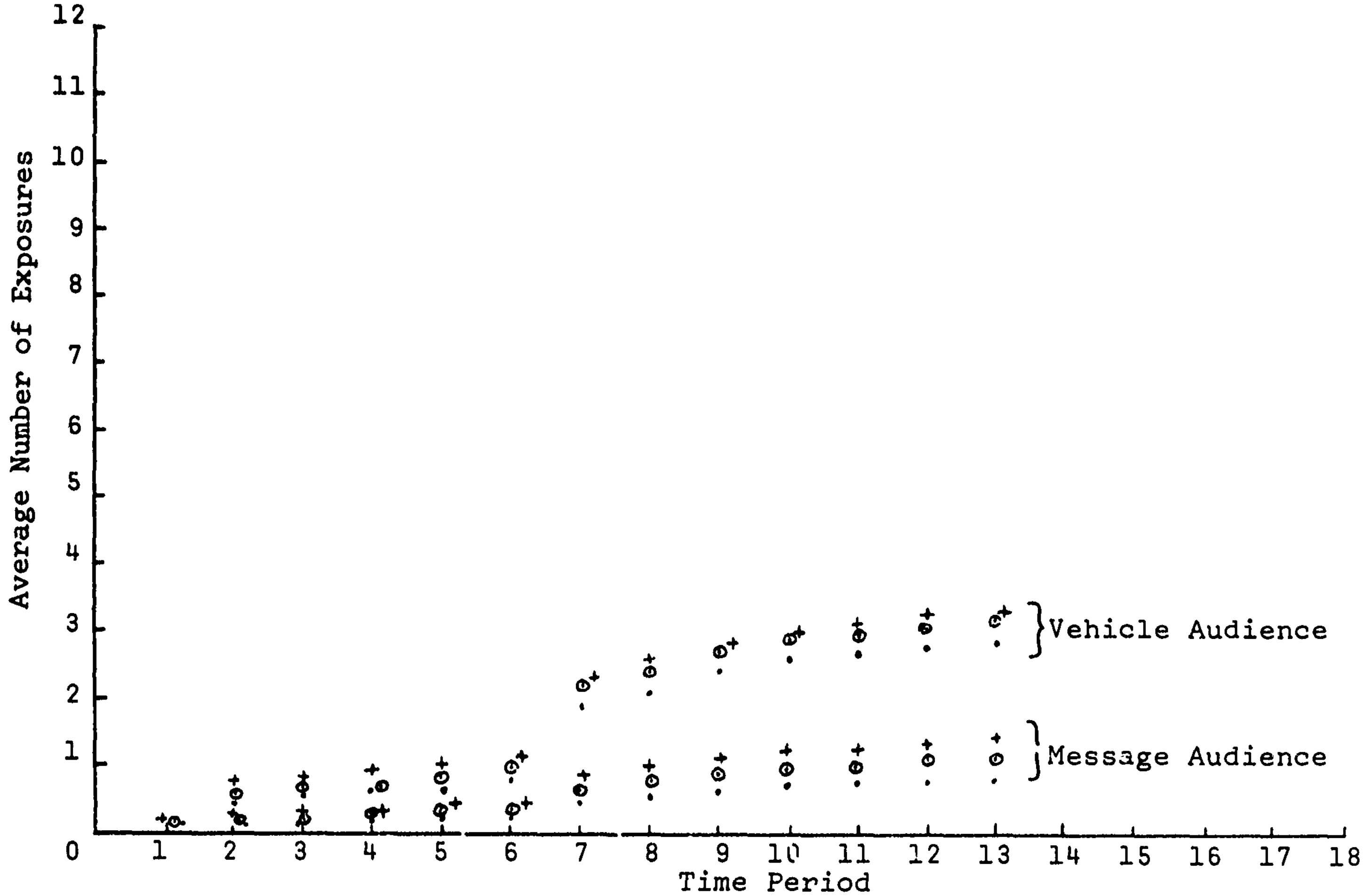


Figure VIII-7. The Average Cumulative Number of Exposures for College (x), High School (⊙), and Grade School (•) Education to the First Trial Scenario

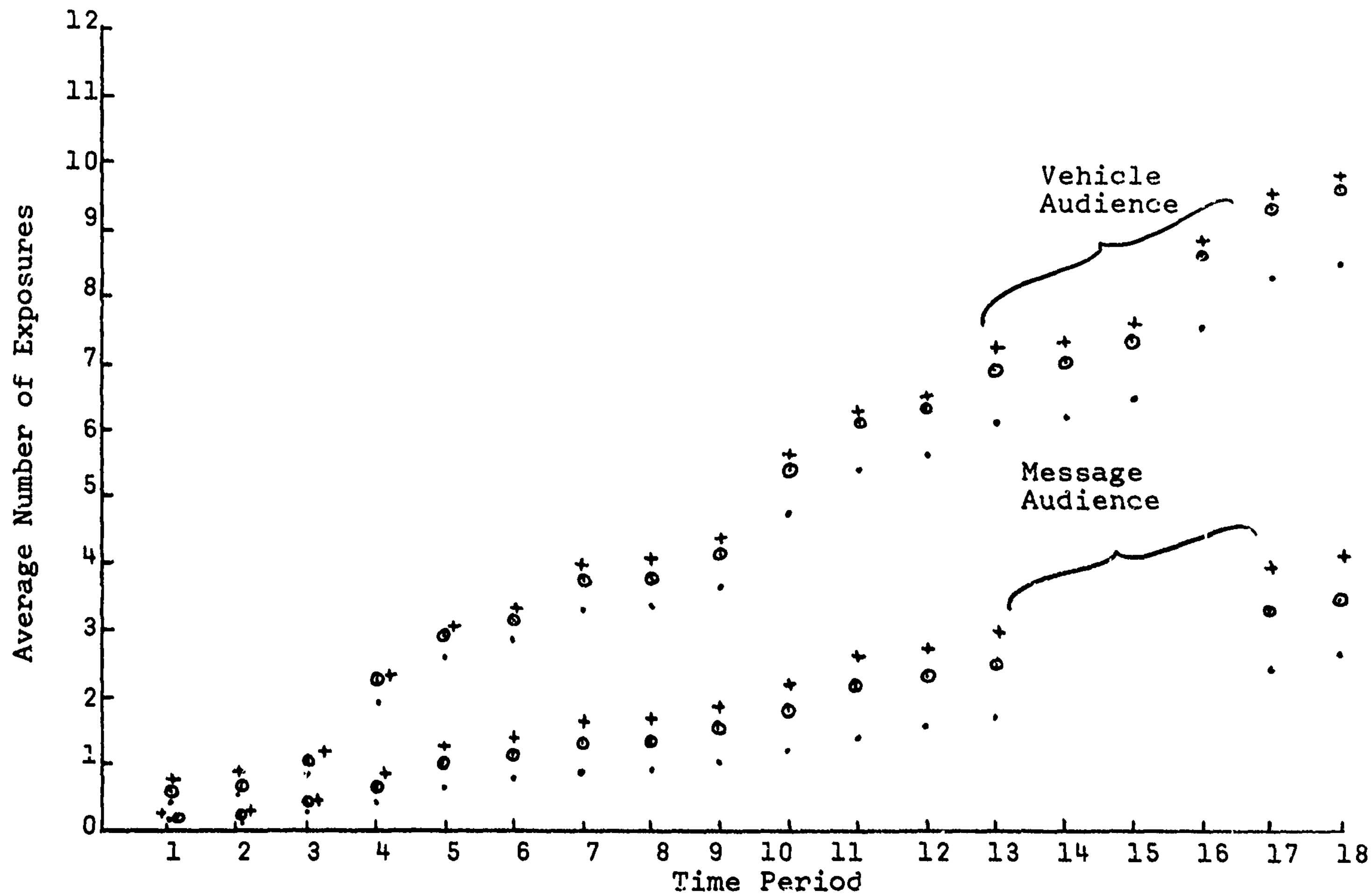


Figure VIII-8. The Average Cumulative Number of Exposures for College (x), High School (⊘), and Grade School (•) Education to the Second Trial Scenario

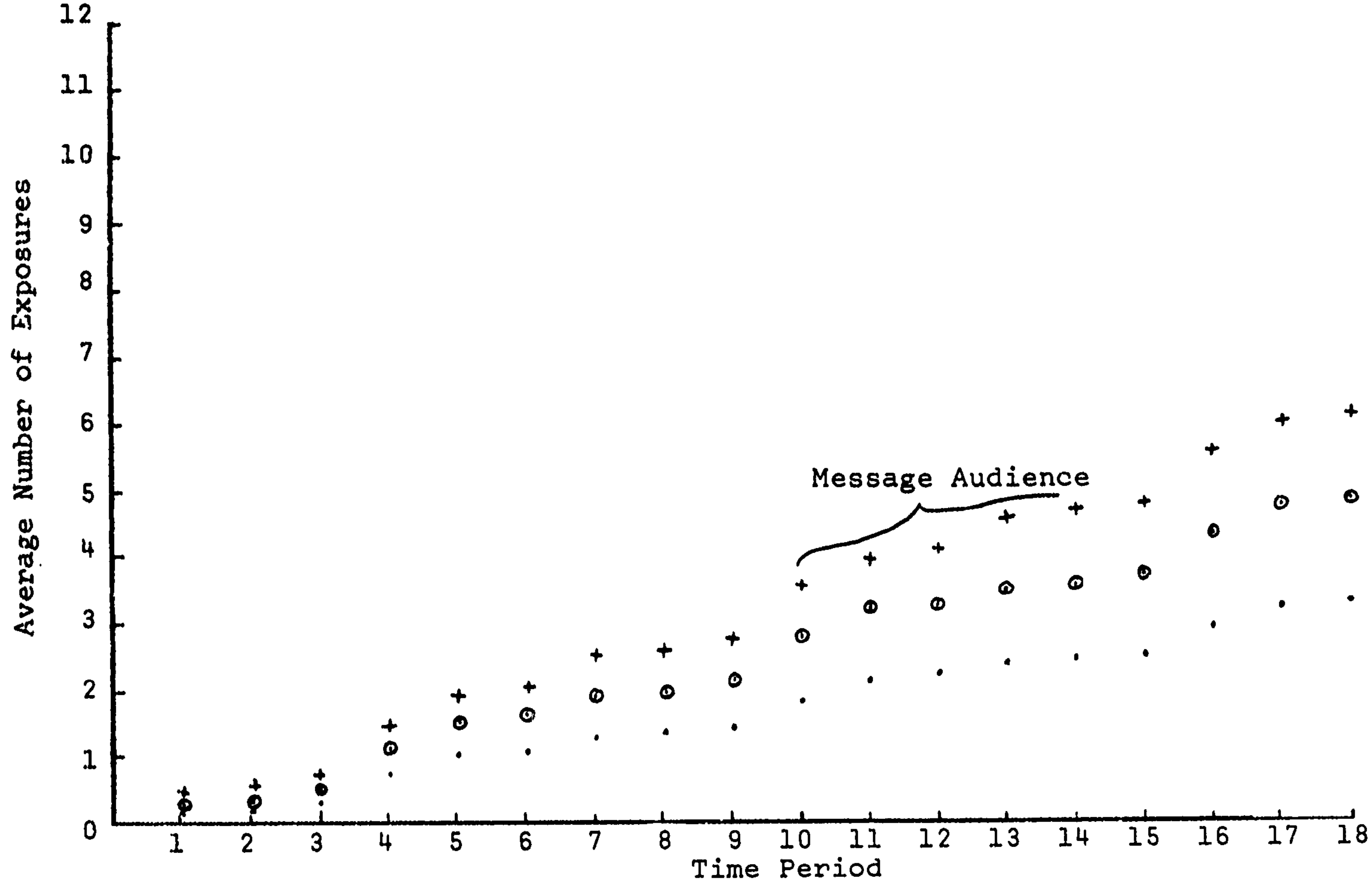


Figure VIII--9. The Average Cumulative Number of Exposures for College (x), High School (◉), and Grade School (•) Education to the Third Trial Scenario

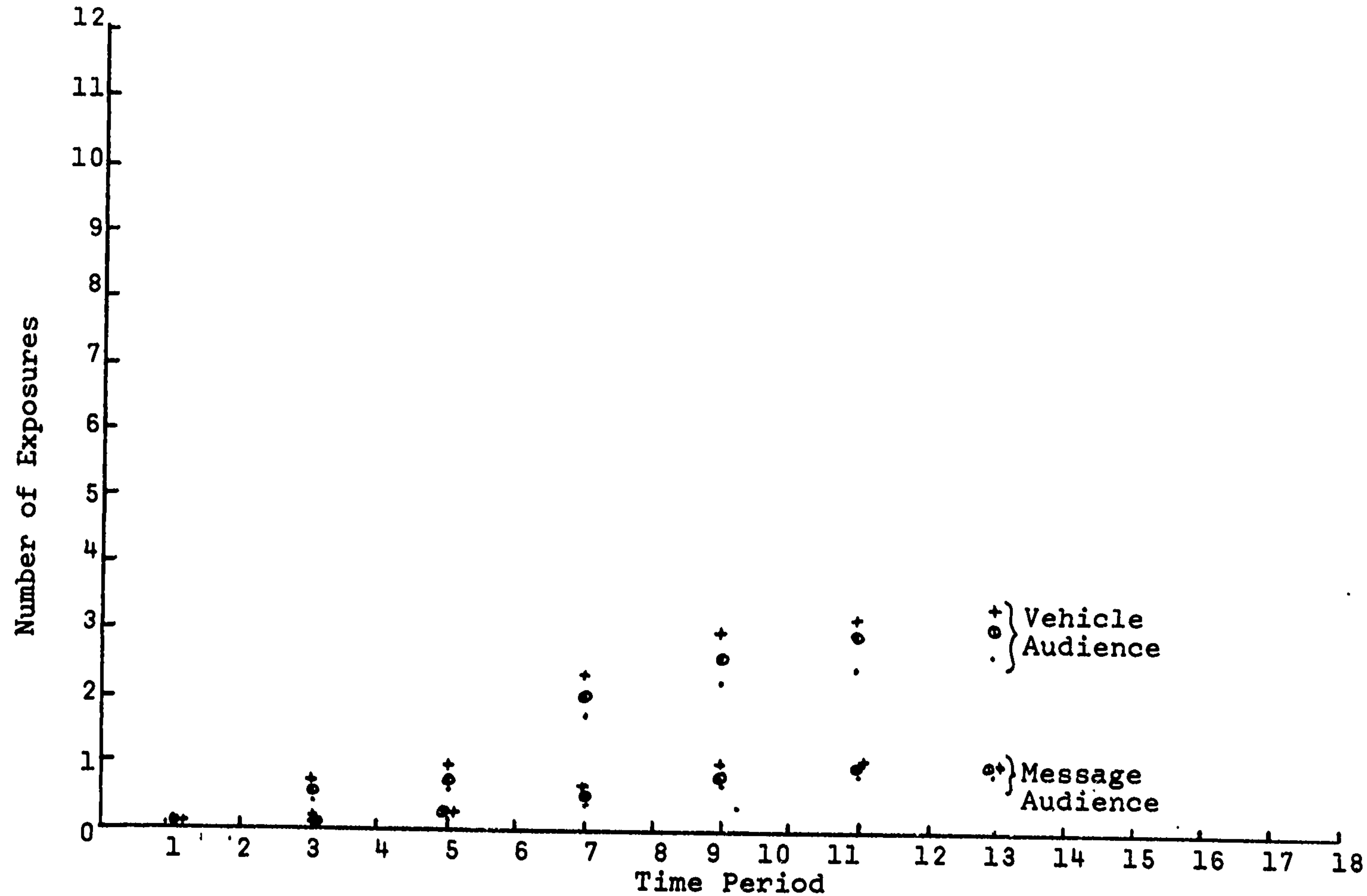


Figure VIII-10. The Average Cumulative Number of Exposures for High (x), Middle (⊕), and Low (•) SES to Theme 1

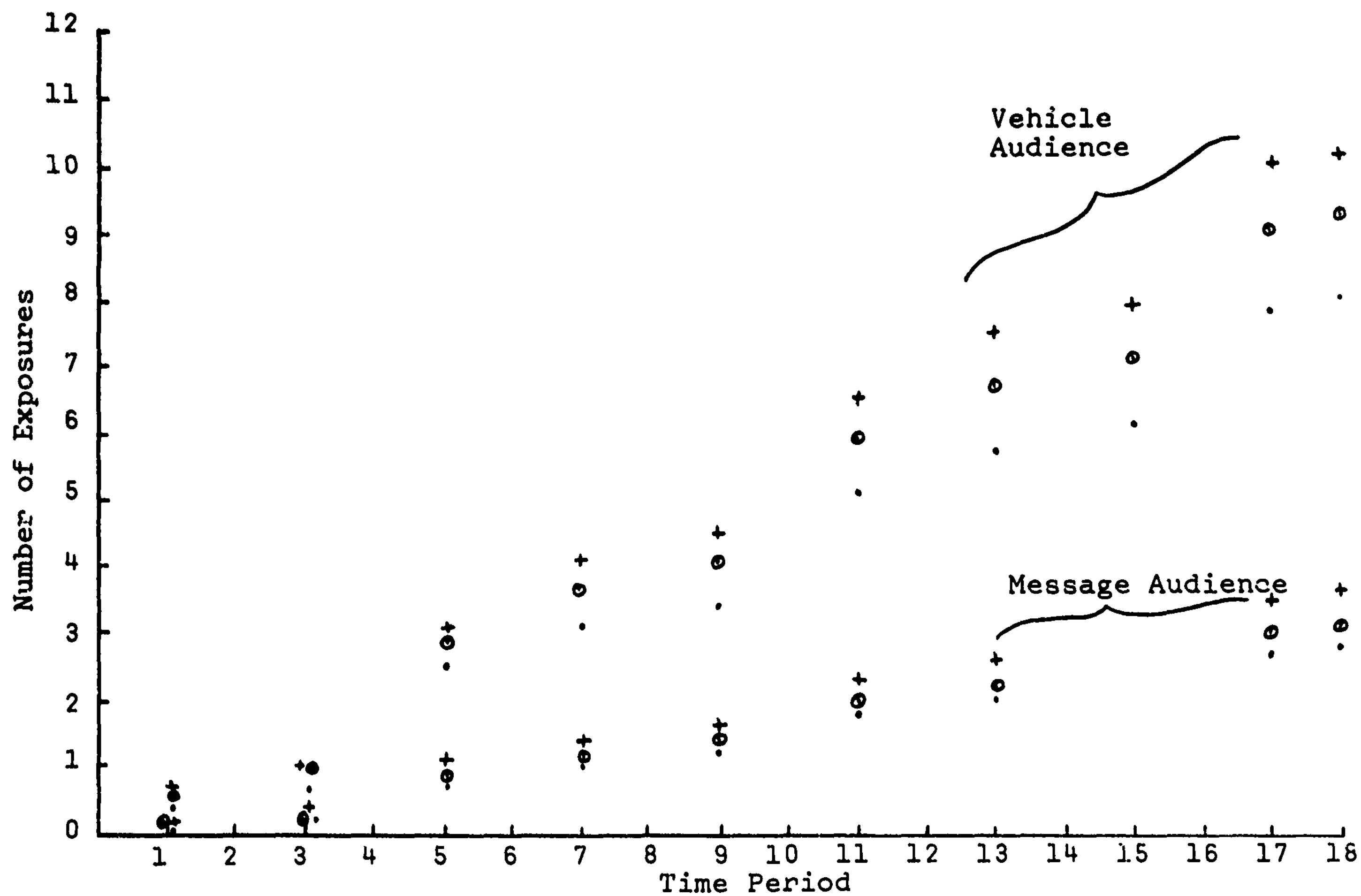


Figure VIII-11. The Average Cumulative Number of Exposures for High (x), Middle (⊙), and Low (•) SES to Theme 2

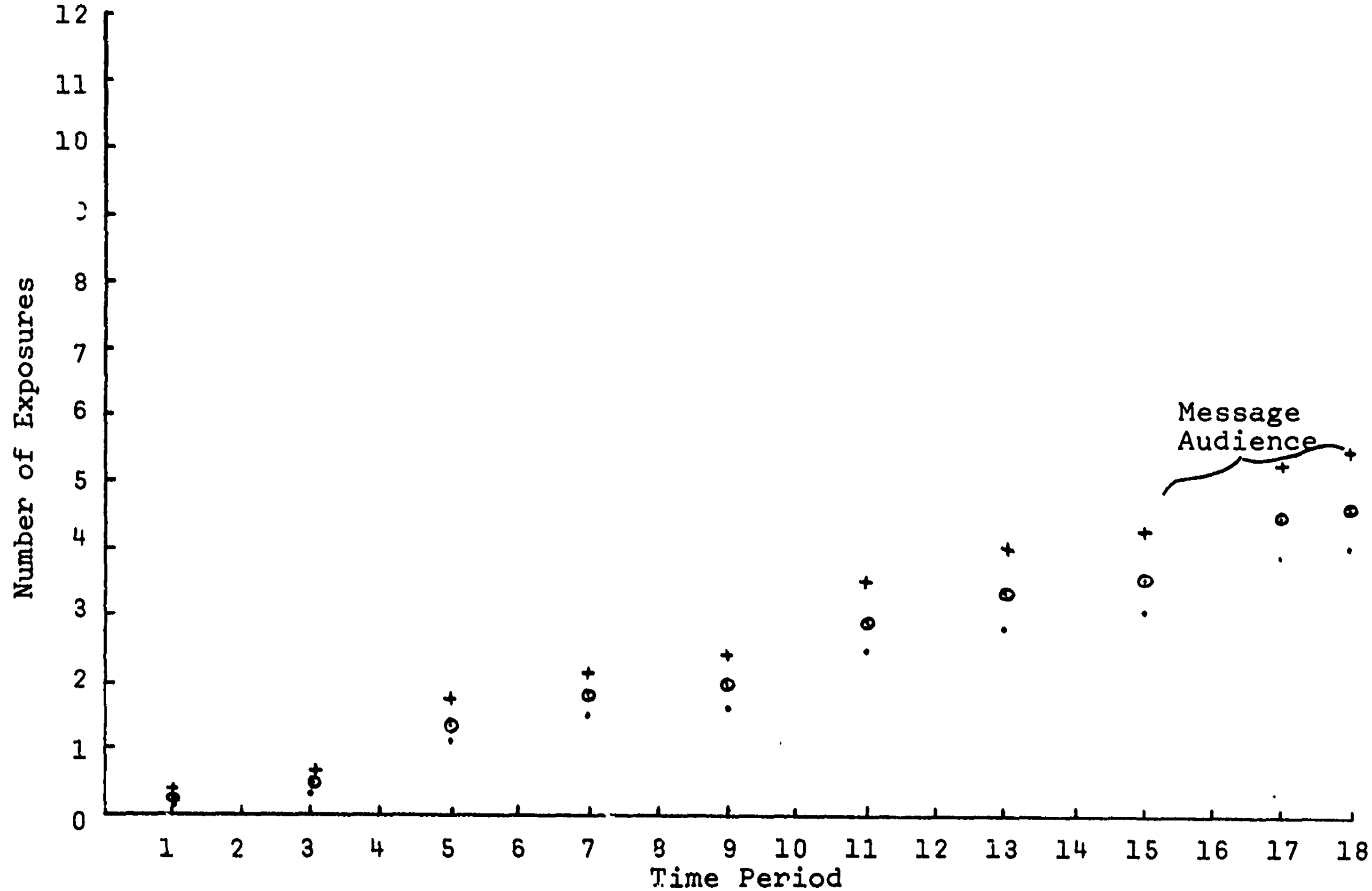


Figure VIII-12. The Average Cumulative Number of Exposures for High (x), Middle (⊙), and Low (•) SES to Theme 3

Table VIII-7. The Average Cumulative Number of Exposures to Vehicles or Messages[a] in Several Population Subgroups for Three Trial Scenarios

Trial[a] Scenario		Average Cumulative Exposure of Population Subgroups					
	Time Period	Total Population	Sex		Age	
			Male	Female	Low	High
1a (One day)	13	3.05	2.82	3.26	3.10	3.02
1b	13	1.06	1.18	.94	1.12	.99
2a (Three day)	18	9.15	8.51	9.77	9.32	9.07
2b	18	3.19	3.54	2.84	3.36	2.98
3a (Large three day)	18	17.39	17.25	17.54	17.28	17.53
3b	18	4.60	5.37	3.84	4.71	4.39

Trial[a] Scenario	Average Cumulative Exposure of Population Subgroups									
	Education			S.E.S.			Interest			
	College	H.S.	G.S.	High	Middle	Low	Low	Low-mid	High-mid	High
1a (One day)	3.24	3.18	2.82	3.39	3.08	2.61	2.93	3.01	2.98	3.31
1b	1.36	1.15	.77	1.21	1.03	.94	.61	.88	1.11	1.61
2a (Three day)	9.72	9.55	8.45	10.16	9.25	8.01	8.78	9.02	8.96	9.95
2b	4.09	3.45	2.31	3.64	3.11	2.84	1.84	2.64	3.34	4.84
3a (Large three day)	19.18	17.86	15.92	19.65	17.60	14.48	15.68	17.20	16.74	19.82
3b	6.13	4.88	3.30	[illegible]	4.49	3.88	2.51	3.77	4.68	7.21

[a]The scenarios designated by the letter (a) are run with message exposure probabilities of 1.0 and can often be interpreted as showing exposure to the vehicles (but see the text above for a caution). The scenarios designated by the letter (b) are run with more realistic message exposure probabilities and show exposure to a message carried in the vehicles.

to one-half of the vehicle audiences regardless of the dimension of the subgroup. For example, in the first theme with a message in each radio broadcast and in each newspaper, there is at the end of the day an average of 3.05 exposures to a vehicle carrying the message and 1.06 exposures to the message itself. The lower figure occurs in the third theme where the ratio of messages in the newspapers to messages in the radio increases: since the message exposure probabilities for newspapers are about one-third of those for radio, we expect and find, that as the proportion of newspaper messages increases, the average number of exposures to the message becomes a smaller proportion of the average number of exposures to the vehicle.

Socio-economic status appears to be the most important variable for the vehicle audience, producing the largest range of exposures, but interest is the most important dimension for the message audience given the distribution of measages and exposure probabilities we have chosen for these trial scenarios. For the first theme, the average number of message exposures at the end of the day ranges from a low of .61 exposures for the lowest interest group to a high of 1.61 exposures for the highest interest group. For the second theme, the figures are just three times as large, reflecting the exposure at the end of three days of messages. Thus the

vehicle audience at the end of three days has experienced an average of approximately 9 exposures and the message audience an average of approximately 3 exposures. For a three-day large news story with three messages in each newspaper, we find a marked increase in vehicle exposure and a moderate increase in the average number of exposures to the message. In this particular case, the values for the vehicle audience are not realistic because the simulation treats each appearance of the message in a vehicle as equivalent to an appearance of the vehicle with one message in it. Thus, the three messages in the morning _Enquirer_ become equivalent to three viewings of the _Enquirer_ and are added to the vehicle audience three times. However, for the message audience the numbers represent the average of the cumulative number of message exposures for each of the subgroups, just as above. In this case, the average number of message exposures over the three days was 4.60 and again, the most important dimension is interest where the average number of exposures ranges from only 2.51 for the lowest interest group to 7.21 for the highest interest group.

In addition to considering each of the population dimensions individually we present in Table VIII-8 those subgroups having the highest and the lowest frequencies of exposure for the combination of the three dimensions sex, socio-economic status, and interest. We find

average message audience differences of five to six hundred percent between the highest and the lowest subgroups. For instance, for the large three-day story, the least exposed subgroup defined by these three dimensions were females of low socio-economic status and low interest who had an average of only 1.14 exposures over the three days. The highest subgroup, on the other hand, were the males of middle socio-economic status and high interest who had almost 7 times the exposure of the lowest group, a figure of 7.78 exposures in the three-day period.

Finally, Table VIII-9 shows the relative importance of the media vehicles in producing message exposures (but not numbers of <u>persons</u> exposed to the messages) to the trial scenarios. For the one-and three-day average stories, the evening news broadcasts are most important followed by the newspapers, and then other news broadcasts at peak listening hours. For the larger story where the newspapers each have three insertions of the message, the newspapers become the most important vehicles, reversing place with the evening news broadcasts, which are now second in order of importance. The more messages carried by the newspapers, of course, the more important become the newspapers relative to the radio broadcasts.

From the analysis of the trial scenarios we conclude that the simulation probably has internal validity, that

Table VIII-8. Population Subgroups by Sex, SES, and Interest, Having the Highest and Lowest Average Cumulative Number of Exposures to Three Trial Scenarios

Trial Scenario[a]	Time Period	Total Population	Most and Least Exposed Subgroups: Most Exposed	Average Exposure	Least Exposed	Average Exposure
1a (one day)	13	3.05	Male, Low SES, High-mid Interest	= 2.25	Female, High SES, High Interest	= 3.80
1b	13	1.06	Female, Low SES, Low Interest	= .30	Female, High SES, High Interest	= 1.74
2a (three day)	18	9.15	Male, Low SES, High-mid Interest	= 6.77	Female, High SES, High Interest	=11.38
2b	18	3.19	Female, Low SES, Low Interest	= .90	Female, High SES, High Interest	= 5.21
3a (Large three day)	18	17.39	Male, Low SES, High-mid Interest	=11.78	Female, High SES, High Interest	=21.43
3b	18	4.60	Female, Low SES, Low Interest	= 1.14	Male, Middle SES, High Interest	= 7.78

[a]The scenarios designated by the letter (a) are run with message exposure probabilities of 1.0 and can often be interpreted as showing exposure to the vehicles (but see the text above for a caution). The scenarios designated by the letter (b) are run with more realistic message exposure probabilities and show exposure to a message carried in the vehicles.

Table VIII-9. Vehicles in the Trial Scenarios Accounting for the Highest Percentage of Total Message Exposures

Trial Scenario	Time Period	Vehicle and Percentage of Message Exposures via the Vehicle
1b and 2b (one-day and three-day stories)	13 and 18	7 PM News, 12.6%; 6 PM News, 11.2%; 8 PM News, 8.8%; Time-Star, 7.9%; Post, 7.5%; Enquirer, 6.4%; 11 PM News, 6.0%; 1 PM News, 5.8%; 8 AM News, 4.6%; 12 AM News, 4.5%; 11 AM News, 4.4%.
3b (large three-day story)	18	Time-Star, 16.6%; Post, 15.7%; Enquirer, 13.3%; 7 PM News, 8.8%; 6 PM News, 7.8%; 8 PM News, 6.2%; 11 PM News, 4.2%; 1 PM News, 4.0%; 8 AM News, 3.2%; 12 AM News, 3.2%; 11 AM News, 3.0%.

it does synthesize consistently the input data, and that it produces plausible expected number of exposures for various population types. For an average story, reported once in each of the vehicles of the Cincinnati mass media system during the day, the average number of exposures at the end of the day is slightly more than one exposure per person. However, this average varies with a range from .30 to 1.74 exposures, depending upon the population type. The most important vehicles seem to be the newspapers, followed by the evening and then the midday and then the morning radio broadcasts, in that order. We turn now to the real scenarios to see what actually happened during the six months of the Cincinnati information campaign.

The Results of the Runs with the Real Scenario

As we noted in Chapter VII above, of our original 17 themes in the content analysis, we ended with a run of 12 themes in the scenario. Four themes were eliminated because of the small number of messages coded in the content analysis for these themes. Unfortunately an important theme concerning control of the atomic bomb was lost because of a programming error in the simulation. In the following pages we shall examine the growth and rates of exposure, the distribution of exposure throughout the

population subgroups, and the important media vehicles, for the twelve scenario themes.

Cumulative Exposure: The Population Totals After Thirteen Time Periods

In Table VIII-10 we summarize some of the relevant input and output statistics for each of the 12 themes actually run in the scenario. The first columns of the table show for each of the themes the number of messages, the number of exposure events resulting from those messages, the average number of exposure events for the population, the number of exposure events per message, the total number of persons exposed to each theme, and the number of persons exposed per message. These statistics are the values at the end of 13 time periods (or the end of the simulated six month interval), except for the twelvth theme which was cut off after 12 time periods. The second column of numbers, the number of exposure events, is the expected number of exposures for the total population. Of course, this average number of exposures is not necessarily the actual number of exposures of any particular individual. The number of exposure events per message is the total number of expected exposures for the population divided by the number of messages. This number depends upon the size of the audience of the vehicle in which the message appears and the

Table VIII-10. The Cumulative Exposure of the Simulation Population After Six Months to the Twelve Scenario Themes

	Number of Messages	Number of Exposure Events	Population Average Number of Exposures	Exposure Events per Message	Number of Persons Exposed (at Least Once)	Number of Persons Exposed (at Least Once) per Message
1. UN Peacekeeping, Promoting Harmony	594	38,042.0	19.0	64.0	1,984.5	3.34
2. US-USSR Hostility	751	45,665.0	22.8	60.8	1,991.9	2.65
3. Violence, Wars, Threats to Peace	1,761	114,418.0	57.2	65.0	1,998.7	1.13
4. Dissension in UN	357	23,544.0	11.8	65.9	1,963.5	5.50
5. Anything about the Veto Power	54	3,763.2	1.9	69.7	1,475.8	27.33
6. Veto only in the Security Council	90	6,098.6	3.1	67.8	1,680.0	18.67
7. Veto Implies Great Power Agreement Necessary	112	7,055.2	3.5	63.0	1,744.6	15.58
8. UN and Human Rights	32	2,201.5	1.1	68.8	1,136.8	35.53
9. Cincinnati Plan Sponsors	142	12,060.0	6.0	84.9	1,842.5	12.98
10. Explicit Explanation of the UN	43	3,400.6	1.7	79.1	1,380.3	32.10
11. Satisfaction with the UN	63	3,815.5	1.9	60.6	1,481.8	23.52
12. Dissatisfaction with the UN	68[a]	4,320.9	2.2	63.5	1,522.4	22.41

[a] Because of a programming error, the twelveth theme was terminated after twelve rather than thirteen time periods or 68 rather than 75 messages.

format of the message, but remarkably this figure is almost constant for many themes. Thus, for every theme except themes nine and ten, the number of exposure events per message lies between 60.8 and 69.7. This indicates that for many themes the distribution of the messages in the vehicles and the distribution of the formating of the messages within each vehicle was similar.

The average number of exposures for a person in the population varies from a low average of 1.1 exposures for the theme of the .N. and human rights, for which there were only 32 messages, to a high average of 57.2 exposures for the most common theme of violence, wars, and threats to peace, which contained 1761 messages. (Recall, however, that we were forced to reduce the number of messages in this theme by one-fifth. Thus the average number of exposures may be more on the order of 71.5 exposures.) We note here that the greatest average number of exposures to themes five through twelve is 5.1 exposures over the six month period. These themes are all themes relating to the United Nations and to the questions of the NORC survey attempting to measure the changes in the Cincinnati population over the six month period. With this very limited number of exposures over the six months and given that those already knowing those facts or having those attitudes about the United Nations which the campaign was attempting to disseminate were

probably more likely to have higher average exposures, it is not surprising that the campaign found little success and little change in the Cincinnati population. Not only do these figures indicate that the messages in the mass media failed to reach those groups which initially were least aware, even if the population had been reached in an undifferentiated way so that everyone had the average exposure, it seems hardly likely that five exposures over six months are enough to cause significant change in the attitudes or information of the population.

Since the number of exposure events per message is relatively constant for most themes, one can estimate, simply by knowledge of the number of messages, the total number of exposures for the population and thus an average number of exposures for each member of the population. We shall see (below) that there are certain regular trends of exposure over the classes of each of the dimensions defining the population types. Might it not be possible to distribute these average number of exposures over the classes of the dimensions according to these trends and then calculate expected values over all the population types to produce the average number of exposure events for each population type? Unfortunately, because of interaction effects, this is not totally

feasible, but might be worth further effort in investigating this possibility.

The number of persons exposed is an expected number of individuals exposed at least once, and of course there is upward bound to this number which is the total population of the simulation. As more messages appear and go out in the mass media, more and more people are reached at least once. People reached in the early time periods accumulate exposures, but do not add to this figure of the expected number of persons exposed at least once. Thus, we see that for those themes with several messages, the expected number of exposures is almost the entire population and even for those themes with very few messages the expected number of exposures is at least half and usually about three-fourths of the population. There is no way of telling how many of these individuals are exposed once or twice in the six month period, and it seems quite unlikely that an exposure of only once or twice in the six month period to most of the messages carried in these themes will make any lasting impression upon the individual. A much more interesting statistic would be, of course, the expected number of persons exposed zero times. one time, two times, three times, and so on, i.e. the frequency distribution of exposures within each subgroup of the population. However, this distribution, although in principle quite easy to calculate,

is impossible to manage in the computer for which the simulation was programmed. The amount of space available for storing statistics was just too limited to allow this kind of detail for this and several other statistics.[6]

As more and more messages appear, the number of times a given individual is exposed is likely to increase; therefore the average number of persons exposed per message decreases with an increasing number of messages and a fixed population size. This is the effect observed in the final two columns of Table VIII-10.

Cumulative Exposure: The Distribution of Exposure over the Population Subgroups

The data above compare total exposure from theme to theme as a function of the number of messages; in addition we can identify for each theme those population subgroups most exposed and those least exposed to each theme. The average cumulative number of exposures for each level of each population dimensions and the overall averages are tabulated in Table VIII-11, at the thirteenth time period. Unfortunately, the data for the first four themes and especially for the third theme have significant errors in their absolute magnitudes.

[6]The simulation is presently being reprogrammed for the IBM 360-65, a much larger machine than the present one, an IBM 7094. On the new machine, the storage problems will be much less critical and these statistics will be available.

Because of a limitation of space in the computer and the programming necessary to take account of this, the data necessary to generate these statistics is accumulated in the computer as the expected number of exposures for each individual in the simulation population via each of the media types for up to six media types. For a simulation population of 2000 and six media types, 12,000 numbers are kept cumulatively up-to-date in order to generate these average number of exposures presented in Tables VIII-11 and VIII-12. In order to keep such large numbers of statistics in the computer the simulation was programmed to pack these statistics four to a word in a computer word. Exposures are then measured in fiftieths allowing as many as 511 fiftieths or 10.2 exposures[7] for each individual via each media type. For those themes which have a large number of messages, the number of exposures for some individuals via some media types exceeded 10.2 exposures. At this point the computer stops accumulating exposures for this person via this media type and the record indicates only 10.2 exposures, thus undercounting exposures for this person. At the time of programming the computer this seemed a logical procedure especially if the output is

---

[7]On the new machine (see Footnote 6, above) this upper limit will be much higher and adjustable by the researcher.

in terms of a distribution of exposures across individuals. Thus, if we know the expected number of individuals exposed zero, one, two, three, four, five, etc., times, then having an upper limit of ten or more exposures is quite suitable for data output in this form; after ten exposures, whatever effect might happen has probably happened. However, for calculating the average number of exposures for subgroups of the population, this procedure gives inaccurate averages, the larger the average number of exposures for the entire population, the more inaccurate are the averages generated by the process.

Each individual can have recorded a maximum of 10.2 exposures for each media type, resulting in a theoretical upper maximum of 61.2 exposures which could be recorded for an individual. Table VIII-10 shows a maximum average of 57.2 exposures for the population for the theme three. Only if the exposures were evenly distributed across the entire population and across each of the media types would the simulation accurately produce the average number of exposures for subgroups of the population. Fortunately, there are other cumulative totals (for the individual media vehicles and each of the six media types) for the number of exposures, and these statistics are accurate. From these we get an estimate of the amount of

Table VIII-11. The Average Cumulative Number of Exposures for Several Simulation Population Subgroups after Six Months

Theme	Total Population	Sex		Age		Education		
		Male	Female	High	Low	College	H.S.	G.S.
1. UN Peace Keeping Promoting Harmony	15.21[a]	18.21	12.65	15.55	14.83	18.28	16.46	12.00
2. US-USSR Hostility	18.08[a]	21.45	15.20	18.37	17.75	22.28	19.15	14.52
3. Violence, Wars, Threats to Peace	29.26[a]	33.36	25.77	28.79	29.78	35.35	30.66	24.32
4. Dissension in UN	10.62[a]	12.88	8.69	10.12	11.08	12.23	11.66	8.43
5. Anything about the Veto Power	1.88	2.32	1.50	1.68	2.06	2.37	2.08	1.37
6. Veto only in the Security Council	3.04	3.75	2.44	2.77	3.29	3.48	3.45	2.27
7. Veto Implies Great Power Agreement Necessary	3.51	4.13	2.99	3.72	3.28	4.09	3.89	2.73
8. UN and Human Rights	1.10	1.43	0.82	1.21	0.98	1.48	1.20	0.77
9. Cincinnati Plan Sponsors	5.11	6.65	3.80	4.47	5.69	6.32	5.54	3.93
10. Explicit Explanation of the UN	1.70	2.25	1.22	1.44	1.93	2.29	1.83	1.22
11. Satisfaction with the UN	1.90	2.38	1.49	2.09	1.69	2.37	2.10	1.39
12. Dissatisfaction[b] with the UN	2.16	2.67	1.73	1.89	2.40	2.73	2.36	1.60

[a]For the first four themes, all the data probably underestimate exposure (especially in the classes of highest exposure) because of the large number of messages and the storage limitation of the computer. See the text for more details.

[b]Because of a programming error, the twelfth theme was terminated after twelve rather than thirteen time periods or 68 rather than 75 messages.

Table VIII-11. (Continued)

Theme	Total Population	SES High	SES Middle	SES Low	Interest Low	Interest Lower-Middle	Interest Upper Middle	Interest High
1. UN Peace Keeping Promoting Harmony	15.21[a]	18.16	15.93	9.81	12.84	14.86	15.50	17.58
2. US-USSR Hostility	18.08[a]	21.42	18.93	11.85	15.26	17.81	18.49	20.69
3. Violence, Wars, Threats to Peace	29.26[a]	33.19	30.28	21.88	25.64	28.99	29.86	32.49
4. Dissension in UN	10.62[a]	12.62	11.22	6.61	8.97	10.36	10.84	12.28
5. Anything about the Veto Power	1.88	2.41	1.98	1.00	1.48	1.82	1.91	2.30
6. Veto only in the Security Council	3.04	3.80	3.19	1.76	2.53	2.94	3.07	3.60
7. Veto Implies Great Power Agreement Necessary	3.51	4.21	3.71	2.15	2.98	3.41	3.59	4.07
8. UN and Human Rights	1.10	1.44	1.16	0.56	0.83	1.05	1.11	1.40
9. Cincinnati Plan Sponsors	5.11	6.51	5.36	2.87	3.85	5.04	5.14	6.37
10. Explicit Explanation of the UN	1.70	2.31	1.77	0.84	1.22	1.64	1.70	2.21
11. Satisfaction with the UN	1.90	2.45	1.99	1.04	1.49	1.85	1.95	2.30
12. Dissatisfaction[b] with the UN	2.16	2.84	2.25	1.17	1.68	2.11	2.21	2.62

[a]For the first four themes, all the data probably underestimate exposure (especially in the classes of highest exposure) because of the large number of messages and the storage limitation of the computer. See the text for more details.

[b]Because of a programming error, the twelfth theme was terminated after twelve rather than thirteen time periods or 68 rather than 75 messages.

undercounting which has taken place for each of the themes due to the large number of messages in the themes. The comparison shows that for theme one, by the thirteenth time period 20 percent of the exposure events were not counted. For theme two the figure is 21 percent, for theme three 49 percent, and for theme four the figure is 9.8 percent. The only other significant undercounting is in theme nine, about 5.5 percent of the exposure events. Of course, this undercounting is not random. It will probably be largest in those cells which have the highest probabilities of exposure. In the data of Tables VIII-11 and VIII-12 the highest average number of exposures for the first four themes should be even higher. In the graphs (below) plotting the growth of exposure for each of the themes, for the first four themes we have extrapolated from the initial time periods when the undercounting is minimal to the correct total number of exposures for the theme in the later time periods.

Noting these problems with the data for the first four themes and especially for the third theme, we turn to the average number of exposures for the classes of each of the population dimensions in Table VIII-11. These average exposures exhibit a wide variation from theme to theme and, for a given theme, within socio-economic classes and educational levels. In every case,

however, we find males are more exposed than females, people with higher education are more exposed than people with lower education, those of higher SES are more exposed than those of lower SES, those with higher interest are more exposed than people of lower levels of interest. Only for the dimension age does the most exposed subgroup vary from theme to theme, but the differences in exposure are so small as to be negligible. We also note that the ratio of the highest average number of exposures to the lowest for each of the dimensions is relatively constant from theme to theme.

Cumulative Exposure: The Most and Least Exposed Subgroups by Sex, by Education, by SES

Table VIII-12 shows the most exposed and the least exposed subgroups by sex, by education, and by SES for each of the themes, and the average number of exposures for each of these subgroups. There are several interesting features to these data. First, we might expect that the most highly exposed group would be college males of high SES; we find instead that it is the high school or grade school educated males of high SES.[8] These findings prompted us to look at the audience

[8]The exceptions to this, the groups of college educated males and females, of low SES, must be discounted since these groups have only two and one members, respectively, and therefore are not large enough to provide any meaningful estimate of exposure.

Table VIII-12. The Most Exposed and Least Exposed Subgroups (by Sex, Education, and SES) for Each of Twelve Themes

	Most Exposed			
Theme	Group	Average Number of Exposures	Group	Average Number of Exposures
1. UN Peace Keeping Promoting Harmony	High SES, male H.S.	29.42	High SES, male, G.S.	27.47
2. US-USSR Hostility	High SES, male, G.S.	30.90	High SES, male, H.S.	29.29
3. Violence, Wars, Threats to Peace	High SES, male, G.S.	40.68	High SES, male, H.S.	39.00
4. Dissension in UN	High SES, male, H.S.	23.11	High SES, male, G.S.	20.87
5. Anything on Veto Power	High SES, male, H.S.	5.44	Low SES, female, Col.	4.44
6. Veto only in Security Council	High SES, male, H.S.	9.55	Low SES, female, Col.	7.10
7. Veto Implies Great Power Agreement Necessary	High SES, male, H.S.	8.13	High SES, male, G.S.	6.66
8. UN and Human Rights	Low SES, male, Col.	3.81	High SES, male, H.S.	3.62
9. Cincinnati Plan (Time Period 12)	High SES, male, H.S.	15.39	High SES, male, G.S.	11.63
10. Explicit Explanation of the UN (Time Period 12)	High SES, male, H.S.	6.00	High SES, male, G.S.	4.34
11. Satisfaction with the UN (Time Period 11)	High SES, male, H.S.	5.84	High SES, male, G.S.	4.23
12. Dissatisfaction with the UN (Time Period 12)	High SES, male, H.S.	6.58	Low SES, female, Col.	5.70

Table VIII-12. (Continued)

Theme	Least Exposed			
	Group	Average Number of Exposures	Group	Average Number of Exposures
1. UN Peace Keeping Promoting Harmony	Low SES, male, G.S.	8.54	Low SES, female, G.S.	4.87
2. US-USSR Hostility	Low SES, male, G.S.	10.44	Low SES, female, G.S.	6.21
3. Violence, Wars, Threats to Peace	Mid SES, female, H.S.	20.84	Low SES, female, G.S.	13.09
4. Dissension in UN	Low SES, male, G.S.	5.40	Low SES, female, G.S.	3.21
5. Anything on Veto Power	Low SES, male, G.S.	0.76	Low SES, female, G.S.	0.46
6. Veto only in Security Council	Low SES, male, G.S.	1.37	Low SES, female, G.S.	0.81
7. Veto Implies Great Power Agreement Necessary	Low SES, male, G.S.	1.67	Low SES, female, G.S.	1.04
8. UN and Human Rights	Low SES, male, G.S.	0.44	Low SES, female, G.S.	0.23
9. Cincinnati Plan (Time Period 12)	Low SES, male, G.S.	2.45	Low SES, female, G.S.	1.25
10. Explicit Explanation of the UN (Time Period 12)	Low SES, male, G.S.	0.69	Low SES, female, G.S.	0.34
11. Satisfaction with the UN (Time Period 11)	Low SES, male, G.S.	0.82	Low SES, female, G.S.	0.47
12. Dissatisfaction with the UN (Time Period 12)	Low SES, male, G.S.	0.91	Low SES, female, G.S.	0.55

Table VIII-13. The Percentages of the Sex by SES by Education Subgroups in the Average Audiences of Each of the Three Weekday Newspapers

	Male			Female		
SES	College	High School	Grade School	College	High School	Grade School
Enquirer						
High	74.1%[a]	22.0%	77.4%	71.7%	55.5%	46.3%
Middle	70.2	55.3	45.4	44.4	36.2	30.3
Low	*[b]	28.3	11.7	*	34.4	20.1
Post						
High	72.2%	74.9%	64.5%	71.7%	59.7%	64.8%
Middle	62.4	52.4	71.2	69.4	49.5	48.0
Low	*	46.0	36.9	*	57.3	38.6
Time-Star						
High	53.2%	39.6%	77.4%	46.7%	52.6%	41.7%
Middle	54.6	64.9	49.5	30.6	59.7	47.3
Low	*	51.3	56.3	*	40.1	44.8

[a]The data are the percent of each subgroup in the audience. Thus 74.1 percent of the college-educated males of high SES were in the audience of the Enquirer.

[b]The asterisks indicate that the number in the population subgroup in the NORC sample was too small to estimate the audiences.

distributions of those vehicles for which the input data specify this distribution by sex, education, and SES. Since the radio audiences were specified only by sex (and were otherwise distributed randomly by the parameter estimation), the vehicles of interest are the newspapers, especially the weekday newspapers.

The individual audiences of these three newspapers as percentages of each subgroup are shown in Table VIII-13. Of the high SES males, those with grade school education seem to have the most coverage (an average of 73.1 percent), followed by the college educated (a 66.5 percent average), and finally the high school educated (a 45.5 percent average). This may explain in part the high exposure of the grade school educated, but, of course, the actual exposure depends also upon the vehicle distribution of the messages, their formating, and the ratios of message exposure probabilities.

A second interesting feature of the table is that the <u>same</u> subgroups tend to be highly and little exposed from theme to theme. Thus, although the magnitude of exposure varies, the relative flow of messages seems not to vary. Since we have not been able to establish different ratios of message exposure probabilities from theme to theme, the only possible way for the distribution of exposures to vary from theme to theme is if the distribution of messages in the vehicles (and thus their

audiences) tends to vary. Evidently they don't. We shall look more closely at this phenomenon when we discuss the correlations of exposures with changes in the panel.

Finally we note the wide range in exposure between the most and least exposed subgroups. Excepting the first four themes where the data must underestimate the highest exposures by about a factor of two, the most exposed groups have nine to fifteen times the exposure of the least exposed group. The dimensions sex, education, and SES do seem to explain much of the variation in exposure.

Cumulative Exposure: The Most Important Media Vehicles

Tables VIII-14 and VIII-15 show the distribution of the number of exposure events for the media types and the most important media vehicles. The exposures by media type indicate that the combination of the afternoon papers was responsible for most of the exposures to the themes in the scenario and the newspapers combined accounted for 35 to 70 percent of the exposures of the population to the themes.

Figures in Table VIII-15 represent the proportion of the total exposures to each theme which are accounted for by each of the six most important media vehicles. For the first four themes involving the important news

Table VIII-14. The Percentage of the Total Exposure Events by Media Type

Theme	Percentage of Exposures to the Theme Via Media Type						
	Afternoon Newspapers (Type 1)	Weekday Morning Newspaper (Type 2)	Sunday Newspaper (Type 3)	Weekday Morning Radio (Type 4)	Weekday Afternoon and Evening Radio (Type 5)	Weekend Radio (Type 6)	Total
1. UN Peacekeeping, Promoting Harmony	45.1%	7.8%	2.7%	8.9%	29.2%	6.2%	99.9%
2. US-USSR Hostility	30.7	10.1	2.5	9.3	40.5	6.7	99.8
3. Violence, Wars, Threats to Peace	35.5	6.6	1.9	6.7	37.3	12.0	100.0
4. Dissension in UN	34.7	5.7	2.7	4.8	42.2	9.3	99.4
5. Anything on Veto Power	26.1	19.7	4.9	6.6	27.3	15.4	100.0
6. Veto only in Security Council	37.6	5.5	4.3	6.3	46.3	0	100.0
7. Veto Implies Great Power Agreement Necessary	30.2	5.1	1.4	6.9	56.3	0	99.0
8. UN and Human Rights	35.3	29.8	14.8	20.1	0	0	100.0
9. Cincinnati Plan Sponsors	63.7	2.6	12.1	0	9.9	11.8	100.1
10. Explicit Explanation of UN	51.4	19.3	17.3	12.0	0	0	100.0
11. Satisfaction with UN	25.9	12.7	4.5	7.2	29.2	20.5	100.0
12. Dissatisfaction with UN	27.4	14.0	3.0	6.2	47.2	2.2	100.0

Table VIII-15. Media Vehicles Accounting for the Largest Percentage of Exposures for each Theme

Theme	Most Important Vehicles and Percentage of Exposures Via the Vehicle (In Order of Importance)					
	1	2	3	4	5	6
1. UN Peacekeeping, Promoting Harmony	Post 29.9%	Time-Star 15.2%	Daily Enquirer 7.8%	7:00 P.M. Weekday News 6.3%	6:00 P.M. Weekday News 5.7%	8:00 P.M. Weekday News 4.4%
2. US-USSR Hostility	Post 15.8	Time-Star 15.0	Daily Enquirer 10.1	7:00 P.M. Weekday News 8.8	6:00 P.M. Weekday News 7.9	8:00 P.M. Weekday News 6.1
3. Violence, Wars, Threats to Peace	Post 18.6	Time-Star 16.9	7:00 P.M. Weekday News 8.2	6:00 P.M. Weekday News 7.1	Daily Enquirer 6.6	8:00 P.M. Weekday News 5.6
4. Dissension in UN	Post 24.7	Time-Star 10.0	7:00 P.M Weekday News 9.4	6:00 P.M. Weekday News 8.4	8:00 P.M. Weekday News 6.5	Daily Enquirer 5.7
5. Anything on Veto Power	Daily Enquirer 19.7	Post 16.7	Time-Star 9.4	7:00 P.M. Weekday News 5.9	6:00 P.M. Weekday News 5.3	Sunday Enquirer 4.9
6. Veto only in Security Countil	Post 28.7	7:00 P.M. Weekday News 10.1	6:00 P.M. Weekday News 9.1	Time-Star 8.8	8:00 P.M. Weekday News 7.0	Daily Enquirer 5.5

Table VIII-15. (Continued)

Theme	Most Important Vehicles and Percentage of Exposures Via the Vehicle (In Order of Importance)					
	1	2	3	4	5	6
7. Veto Implies Great Power Agreement Necessary	Post 19.3	7:00 P.M. Weekday News 12.4	6:00 P.M. Weekday News 11.1	Time-Star 10.9	8:00 P.M. Weekday News 8.6	11:00 P.M. Weekday News 5.9
8. UN and Human Rights	Daily Enquirer 29.8	Post 29.2	Sunday Enquirer 14.8	12:00 A.M. Weekday News 6.2	Time-Star 6.1	8:00 A.M. Weekday News 4.9
9. Cincinnati Plan Sponsors	Time-Star 34.9	Post 28.8	Sunday Enquirer 12.1	Daily Enquirer 2.6	7:00 P.M. Saturday News 2.3	6:00 P.M. Saturday News 2.2
10. Explicit Explanation of UN	Time-Star 31.2	Post 20.2	Daily Enquirer 19.3	Sunday Enquirer 17.2	8:00 A.M. Weekday News 2.8	12:00 A.M. Weekday News 2.7
11. Satisfaction with UN	Time-Star 15.5	Daily Enquirer 12.7	Post 10.3	7:00 P.M. Weekday News 6.3	6:00 P.M. Weekday News 5.7	Sunday Enquirer 4.5
12. Dissatisfaction with UN	Time-Star 23.3	Daily Enquirer 14.0	7:00 P.M. Weekday News 10.2	6:00 P.M. Weekday News 9.2	8:00 P.M. Weekday News 7.2	11:00 P.M. Weekday News 4.9

stories of the period--matters of peace and war, great power conflicts, dissension in the United Nations, etc.,--the vehicle carrying the largest proportion of exposures was always the newspaper, the Post. For both themes one and four the Post is the most important by far of the media vehicles; in the other two it is followed closely by the Time-Star. The average audiences of the three newspapers (for the Time-Star 50.2 percent of the population, for the Post 47.6 percent, and for the Enquirer 39.2 percent) indicate that the increased number of exposures of the Post over the Time-Star cannot be accounted for simply by the differences in newspaper audiences but by a combination of number of stories carried and the formating of these stories in the newspaper. For these important news stories the Time-Star was the second most important vehicle, often followed by the Enquirer.[9] Usually the weekday evening news broadcasts at six, seven, and eight o'clock were additional important vehicles in exposing the population of these news stories.

For several of the themes with low message frequencies and more specialized kinds of messages, e.g.

---

[9]In conversations with the author, several members of the publicity committee for the information campaign remarked that, among the three newspapers, they had the most success in placing news items in the Post. Also, at that time the Enquirer was probably the most liberal, the Time-Star the least liberal, of the newspapers.

explanation of the United Nations, messages about the Cincinnati Plan sponsors, or the United Nations and its relationship to human rights (themes eight, nine and ten), the Sunday newspaper was relatively more important reflecting the special character of these news features. Also, for the ninth theme, the Saturday evening news broadcasts were important.

In one sense, however, these comparisons of media vehicles were misleading in that a weekday news vehicle accumulates average exposures over six days of each week while the weekend news vehicles appear only once in each week. Perhaps a more legitimate comparison would be made by comparing one-sixth of the weekday exposures and one-fifth of the weekday news broadcasts to the exposures of the weekend vehicles.

Cumulative Exposure: The Duplication of Exposure Across Themes

The final table, Table VIII-16 shows the expected audience duplication between each of the first three themes and the other themes by sex, by education, and by socio-economic status. As originally conceived this kind of statistic would allow one to explore the clustering of exposures to the various themes and therefore perhaps explain some of the interrelatedness of the attitudes or information within the population. It will be recalled that because of the limitations of space in

the computer the duplication statistics were a compromise; we might like to know the audience common to any possible combination of themes in the simulation, by any of the dimensions of the populations, for any time period. Unfortunately, there is not enough capacity in the present machine to maintain all of these statistics. In fact, there is not enough capacity in the machine to calculate duplications between every pair of themes and so we have settled upon the compromise of allowing duplication statistics between any theme and each of the first three themes run in the simulation. Therefore one must decide in advance which particular duplications are of greatest interest and make those themes which are of central importance the first three themes of the simulation.

In addition, there is another problem with the statistics generated by the duplication calculations. The duplication is calculated in the following manner: for each individual the cumulative probability of not having been exposed at all to each theme is maintained (or some variant of this that allows the calculation of this probability) for each individual. From these non-exposure probabilities, we calculate for each theme the probability of at least one exposure, the product of these probabilities for any pair of themes is the probability of exposure to both of the themes. The sum

Table VIII-16. The Expected Audience Duplication Between the First Three Themes and Each of the Other Themes by Sex, Education, and SES

Theme	Proportion of Each Group in the Audience (at Least Once) of Both the Given Theme and the First Theme (or the Second or the Third Theme)[a]							
	Sex		Education			SES		
	Male	Female	Col.	H.S.	G.S.	High	Middle	Low
1. UN Peacekeeping Promoting Harmony	-	-	-	-	-	-	-	-
2. US-USSR Hostility	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.97
3. Violence, Wars, Threats to Peace	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99
4. Dissension in UN	1.00	0.97	1.00	0.99	0.96	1.00	0.99	0.94
5. Anything on Veto Power	0.81	0.68	0.84	0.78	0.64	0.84	0.76	0.55
6. Veto only in Security Council	0.90	0.79	0.91	0.87	0.76	0.90	0.86	0.70
7. Veto Implies Great Power Agreement Necessary	0.92	0.82	0.93	0.89	0.79	0.92	0.88	0.74
8. UN and Human Rights	0.66	0.50	0.69	0.61	0.46	0.68	0.59	0.38
9. Cincinnati Plan Sponsors	0.96	0.86	0.98	0.93	0.84	0.96	0.92	0.80
10. Explicit Explanation of UN	0.79	0.61	0.83	0.72	0.58	0.81	0.71	0.49
11. Satisfaction with UN	0.82	0.67	0.85	0.78	0.64	0.84	0.76	0.56
12. Dissatisfaction with UN	0.75	0.62	0.79	0.71	0.58	0.79	0.70	0.49

[a]Because the proportion of each group in the audience (at least once) of each of the first three themes was about 0.99, the duplication of audience with any other theme is almost exactly the same for each of the first two themes. For this reason only one proportion is entered in each line of the table.

of these probabilities of exposure for a population subgroup or cell is the expected number of people in the subgroup exposed at least once to each of the pair of themes. Of course, this expected number includes people who are exposed only once to each theme or once to one theme and twice to another theme, etc. As we have argued above, it would be preferable to have the statistic which represented the expected number of people exposed to each of the themes at least some minimal number of exposures. If five, or six, or seven exposures to a theme are necessary for any significant change over six months for most of the kinds of messages and themes in the present simulation, then we would prefer to know the expected number of people exposed at least five times to each of the two themes.

Since the statistic calculated in the present simulation is the proportion of each subgroup exposed at least once to one of any of the first three themes and any one of the other themes, and since the percentage of any subgroup in any of the first three themes who are exposed at least once is extremely high, approximating one hundred percent (Table VIII-10 shows 1984.5 of a total of 2000 persons exposed at least once to the first theme), then the statistics presented in this table represent almost exclusively the proportion in the subgroups exposed at least once to themes other

than themes one, two, or three. In fact these statistics are almost identical to the percentage of each audience type exposed for each of the other themes. This makes the table of duplications rather uninteresting since it simply reflects the exposure to themes other than the first three themes. Had we anticipated this result, we could have run other themes as the first three themes, and gotten at least some duplication statistics between the less frequently occuring themes.

There are few surprises in these data. Exposure is distributed across the dimensions of the population in just about the way which one might expect, except that there were some interactions when we put these dimensions together, probably due to the proportionally larger newspaper vehicle audiences in the high school and grade school educated males. The weekday newspapers and the morning, mid-day and evening news broadcasts are the most important vehicles, in terms of number of exposures. The very useful fact about these data is that they provide actual number of exposures for each population subgroup to the themes in the media system and trace the growth of these exposures over time periods. These numerical data can be incorporated into a model of some effects of exposure to the media system. Following the graphs of the growth of cumulative exposure, we will propose two such models and examine the

resulting correlations between exposure to the various themes and changes in information, attitudes and opinions in the population over the six months.[10] First, however, we show the growth of exposure for several dimensions over the thirteen two-week time periods for the twelve themes of the scenario.

The Growth of Cumulative Exposure

The first three graphs displayed below show, for the first theme, the growth of the percentage exposed at least once, by sex, education, and SES. As we noted above in the discussion of Table VIII-10, nearly everyone is exposed at least once by the thirteenth time period for all of the themes. In the early time periods, there are small differences between subgroups in the cumulative percent exposed, but these differences disappear quite quickly. Since these statistics are not as interesting as the cumulative <u>number</u> of exposures and since most of the curves for most of the themes grow to greater than ninety percent exposed, we turn

[10]We note that the simulation is not presently programmed in the best way to easily attach a model of change or conversion in the population or the spreading of information by word of mouth resulting from the exposures to the messages. Because of the limited capacity of the machine, we can not store and maintain the distribution of numbers of people with different numbers of exposures for each population subgroup. However, in the reprogrammed version, these data will be stored and available at a later time.

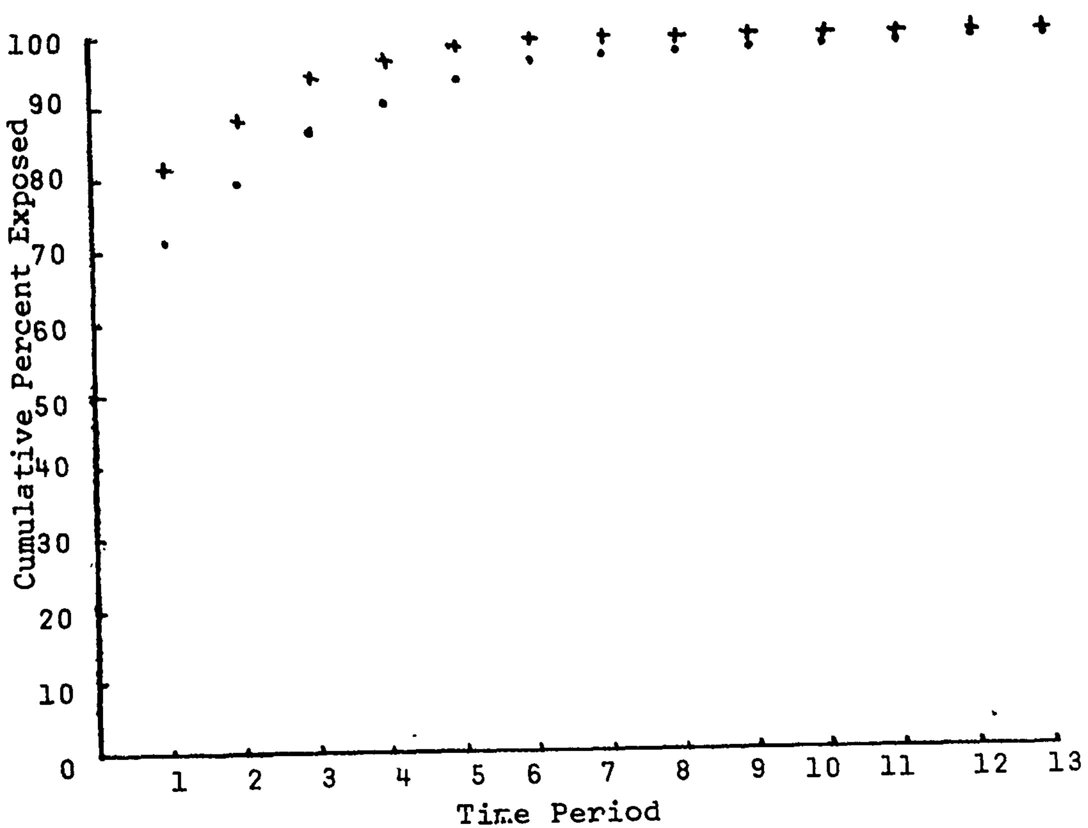


Figure VIII-13. The Cumulative Percentage of Males (+) and Females (•) Exposed at Least Once to the First Theme

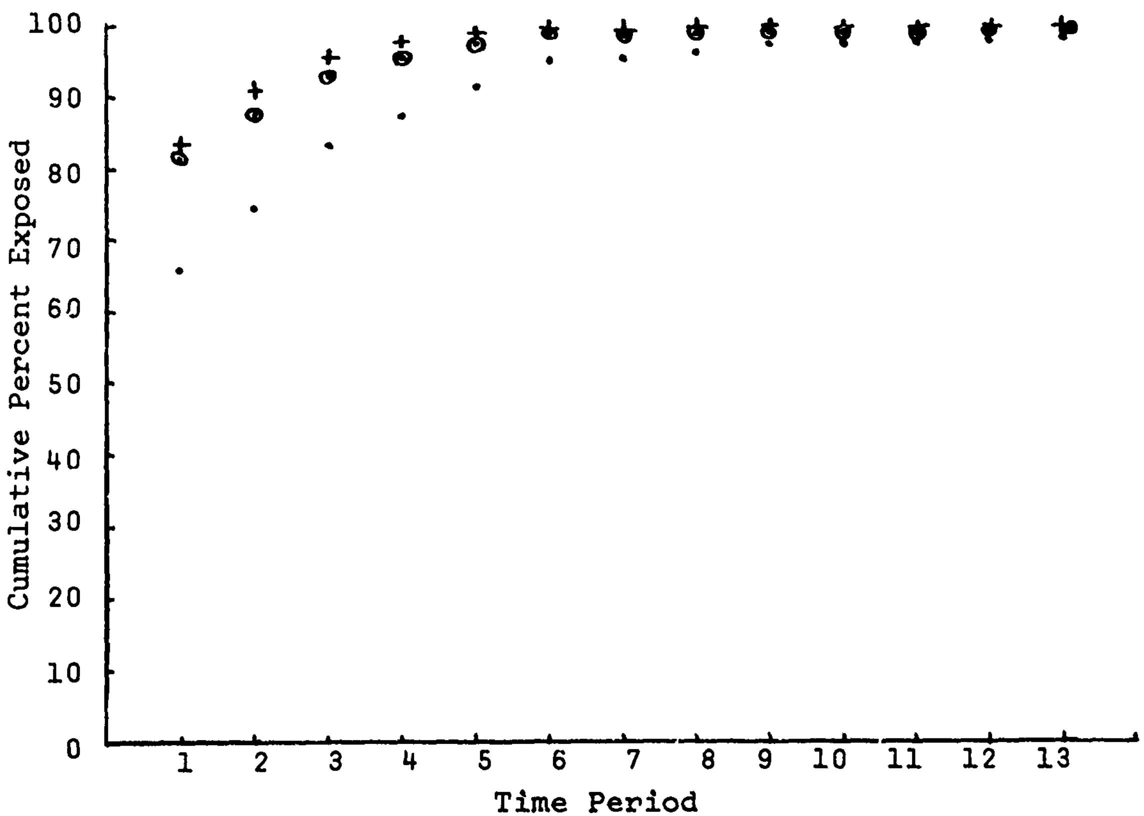


Figure VIII-14. The Cumulative Percentage of College (+), High School (·), and Grade School (.) Educated People Exposed at Least Once to the First Theme

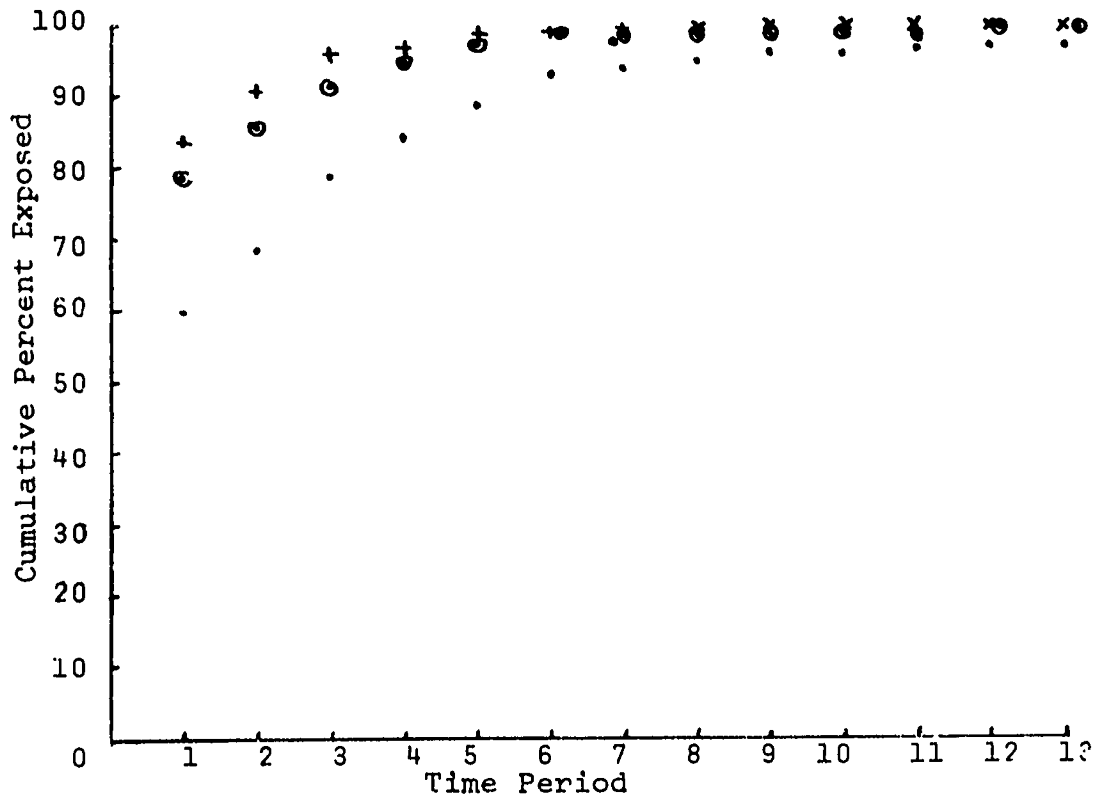


Figure VIII-15. The Cumulative Percentage of High(x), Middle (●), and Low (•) SES People Fxposed at Least Once to the First Theme

now to the graphs of the growth of the cumulative number of exposures.

The Cumulative Exposures of Several Population Subgroups to Each of the Simulation Themes

We have noted above that the record of exposure by population type for each of the first four themes is inaccurate and underestimates the true number of exposures because of the computer storage capacity problem and the high number of messages for each of these themes. However, from the record of exposures either by media or media types, we have accurate values for the expected number of exposures for the entire population, and therefore for the average number of exposures for the population as a whole. If we look at exposures for any exhaustive set of subgroups, e.g., males and females or males and females of college, high school, and grade school education, then the true average exposures of these subgroups of the simulation population must distribute themselves about the population average value.

In the graphs below showing the cumulative number of exposures for various subgroups of the population, we have for the first four themes attempted to correct for the undercounting. Let us take as an example the case of the third theme (the worst case) in which at

the end of thirteen time periods the number of exposures counted in the subgroups was only fifty percent of the actual total number of exposures. In the first time periods, before the number of messages grows large, the error in cumulating exposures is small; however, by the fifth time period the error equals 8.3 percent, i.e., the exposures in the subgroups underestimate the true exposures by about 8.3 percent. Recalling the discussion of the source of this error above (p. 343) it seems likely that in any set of population subgroups, the smaller the average number of exposures actually counted, the less the percent of error in that count. Thus, for the graph and growth of exposure by sex for the third theme, since the females are the least exposed of the two groups, it is probable that the degree of underestimation of their exposure is less than that of the males. Since the amount of underexposure as a whole at the fifth time period is 8.3 percent, we might estimate that for the females the amount of underexposure is on the order of 5 percent compared with perhaps 10 or 11 percent for the males. Taking the value of 5 percent for the females we calculate a revised estimate of the number of exposures. From the total number of exposures for the population at the fifth time period and the number of exposures for the females, the difference is the number of exposures for the males, from

which the average is easily calculated. We have proceeded in this fashion for each time period, producing more realistic values of exposures for the various subgroups of the population.

For each theme we have plotted the growth of exposure by sex, education, and socio-economic status. The six month interval is divided into thirteen two-week time periods, beginning approximately September 15, 1947 and ending March 15, 1948. Thus, the end of the first simulated time period occurs on September 29, 1947.

During the six month period the most important international news events (the events most in the newspaper headlines) concerned the near war between Jews and Arabs in Palestine and partitioning of the area. There were constant bombings, rioting, and violent encounters between the Jews, Arabs, and the British who nominally were in control. On November 11, 1947, the United States and Russia came to an agreement in the United Nations to partition Palestine. On November 29, the United Nations General Assimbly approved the partition and the Arab states walked out, declaring that the partitioning meant war. From this point on, the violence increased markedly with both Jewish and Arab mobs roaming the streets of Jerusalem, burning and looting. By February, the United Nations was attempting

to raise a force to replace the British in Palestine.

Palestine, however, was not the only tumultuous spot on the globe. In December, communist-led strikes of dock and utility workers in France and Italy halted most transportation and commerce. Italians and Yugoslavs battled over elections in Trieste. During January, 1948, a communist government and guerrilla movement was formed in Greece and attacked towns near the Albanian border. The United States responded to both of these situations by dispatching 1000 marines to waters near Trieste and Greece. In the far east, United States and Russian armies faced each other in Korea, and Pakistan sent 4000 troops against Indians in Kashmir.

Also, in January, the United States succeeded in forming a United Nations "little assembly" which was to remain in session throughout the year; however, Russia and the satellites boycotted it. In early February, Russia called for debate in the United Nations on the "slave labor" Taft-Hartley law.

These, then, were the events which produced the messages measured in the content analysis. The graphs of exposure by time period follow.

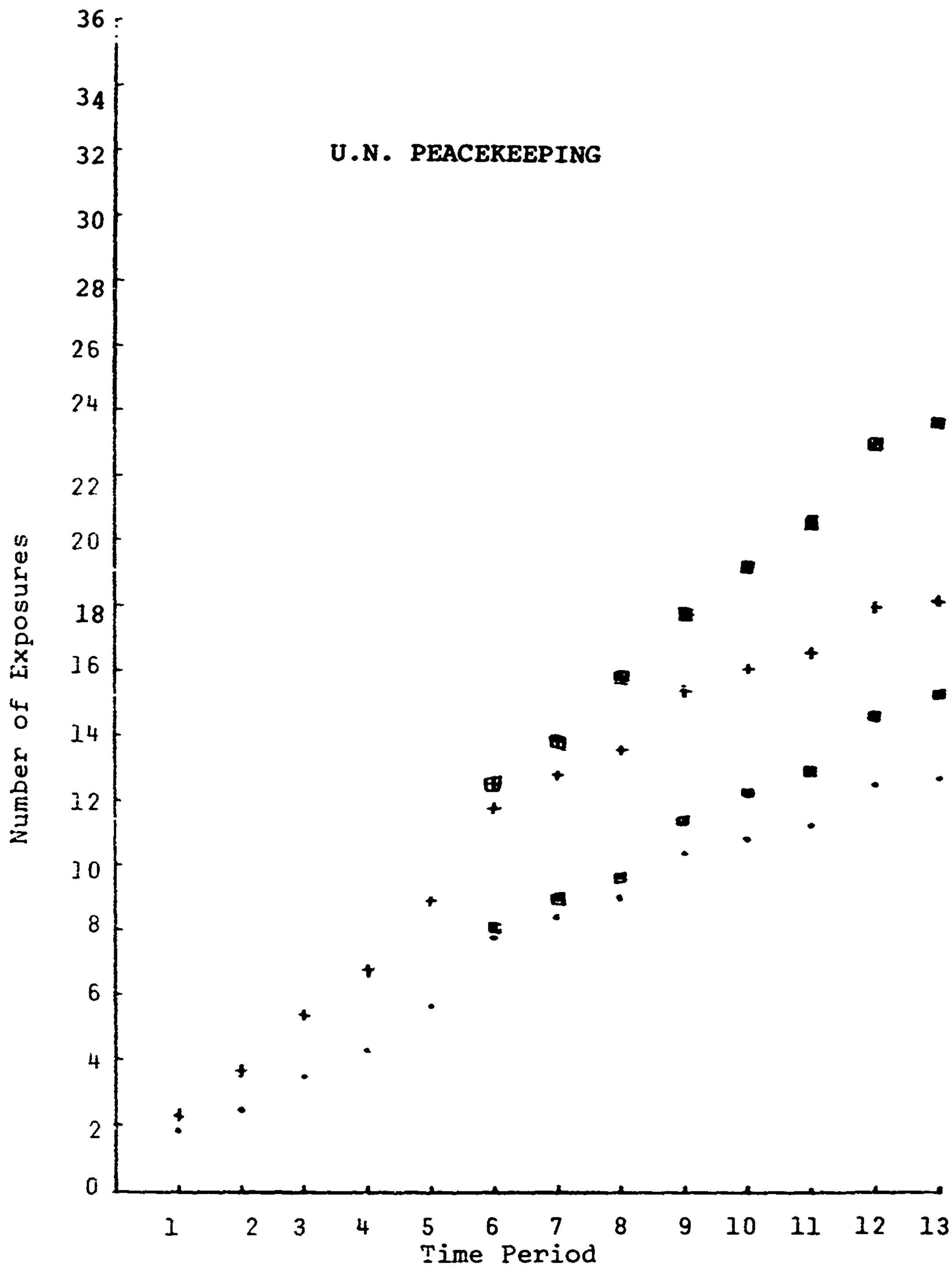


Figure VIII-16. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 1

(points in boxes are adjusted figures)

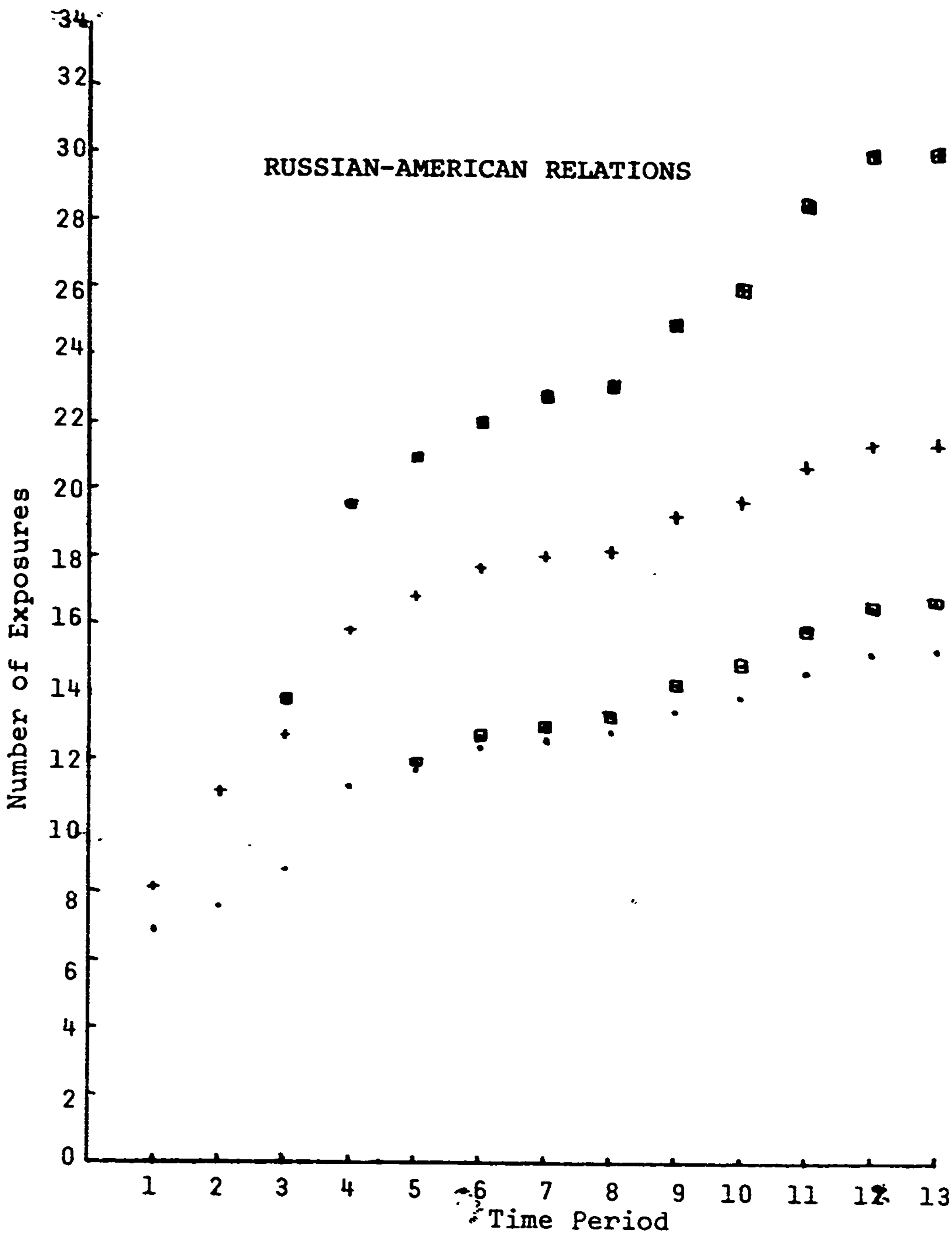


Figure VIII-17. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 2

(points in boxes are adjusted figures)

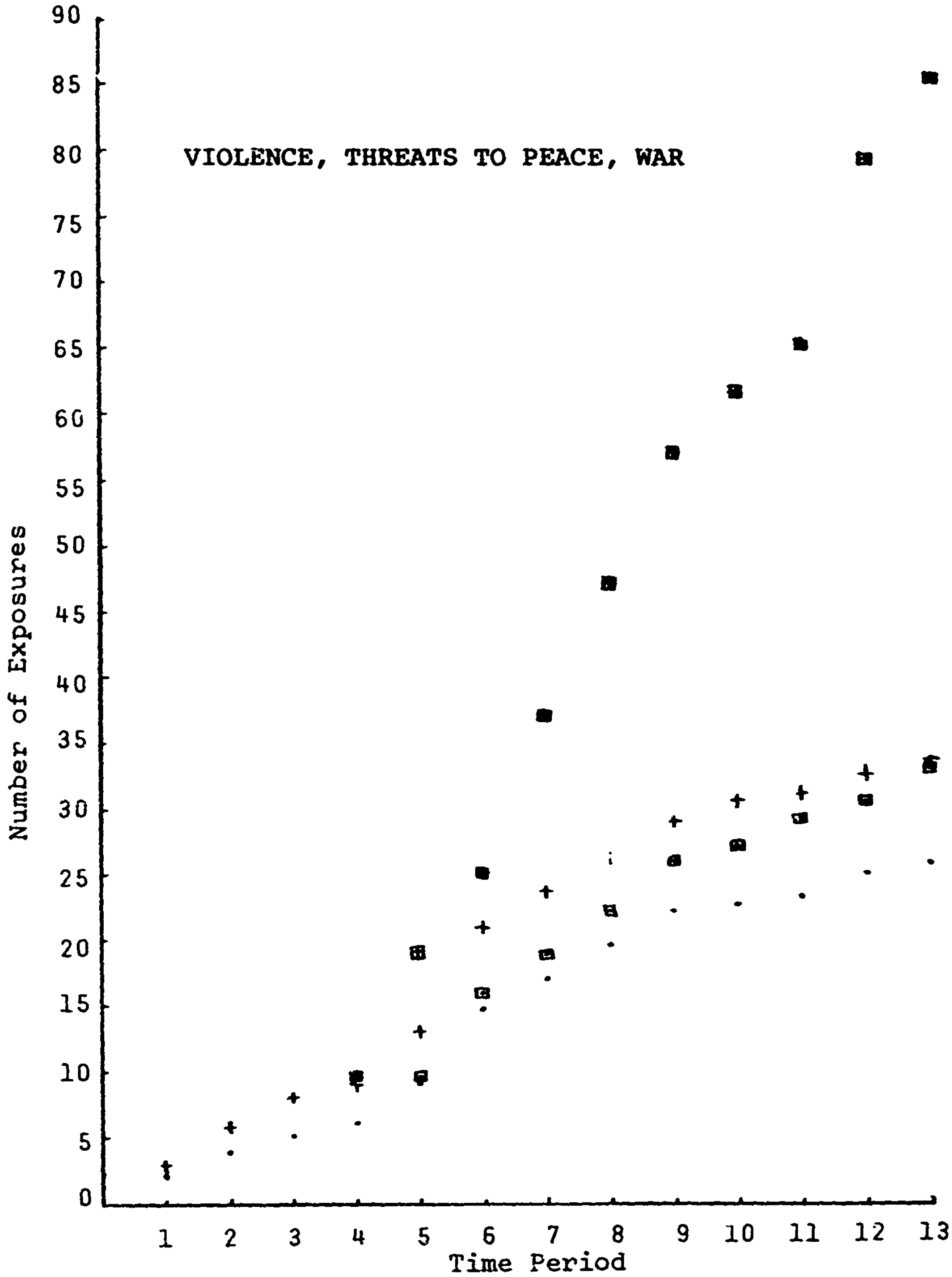


Figure VIII-18: The Average Cumulative Number of Exposure Events for Males (+) and Females (•) to Theme 3

(points in boxes are adjusted figures)

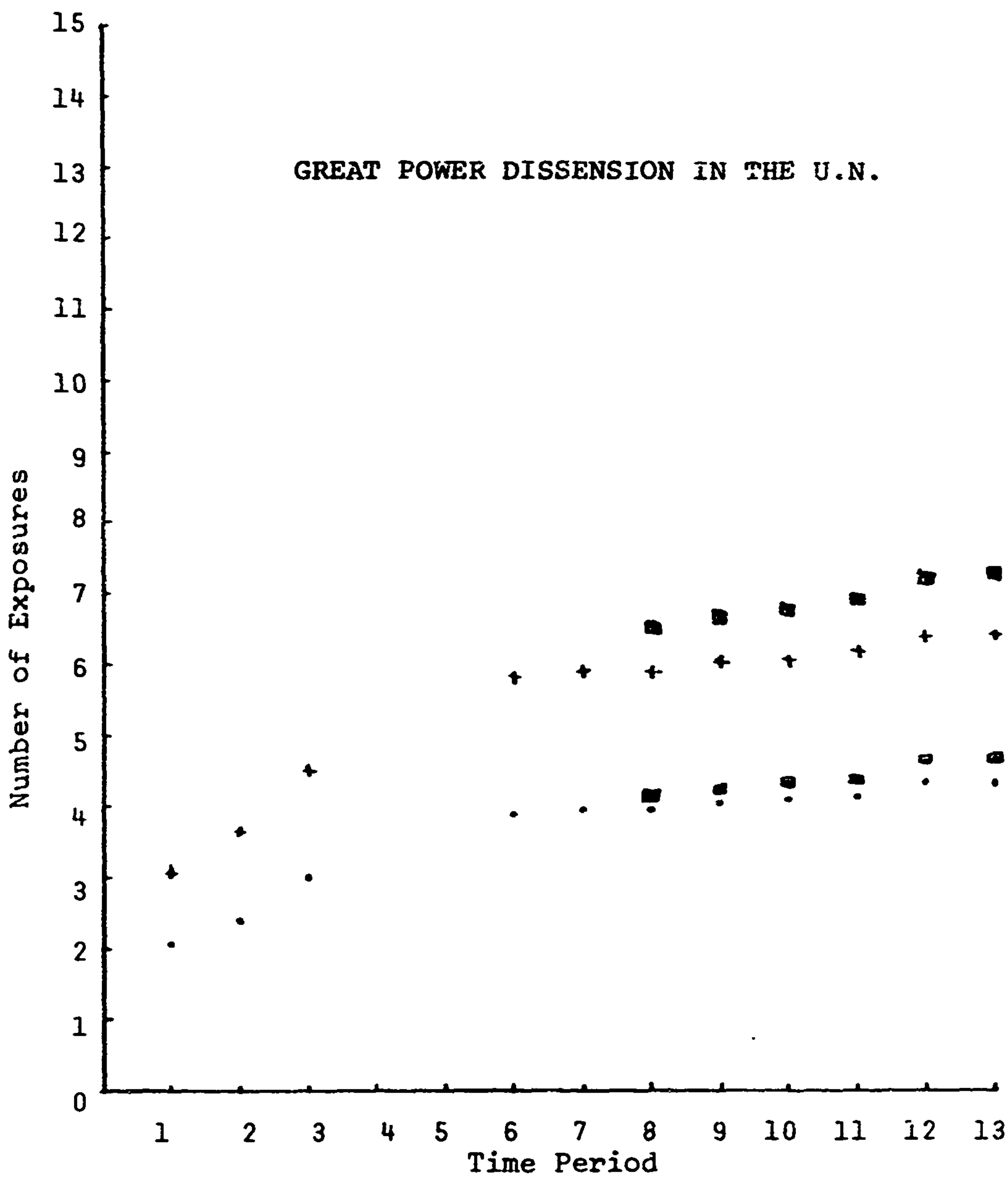


Figure VIII-19. The Average Cumulative Number of Exposures For Males (+) and Females (•) to Theme 4

(points in boxes are adjusted figures)

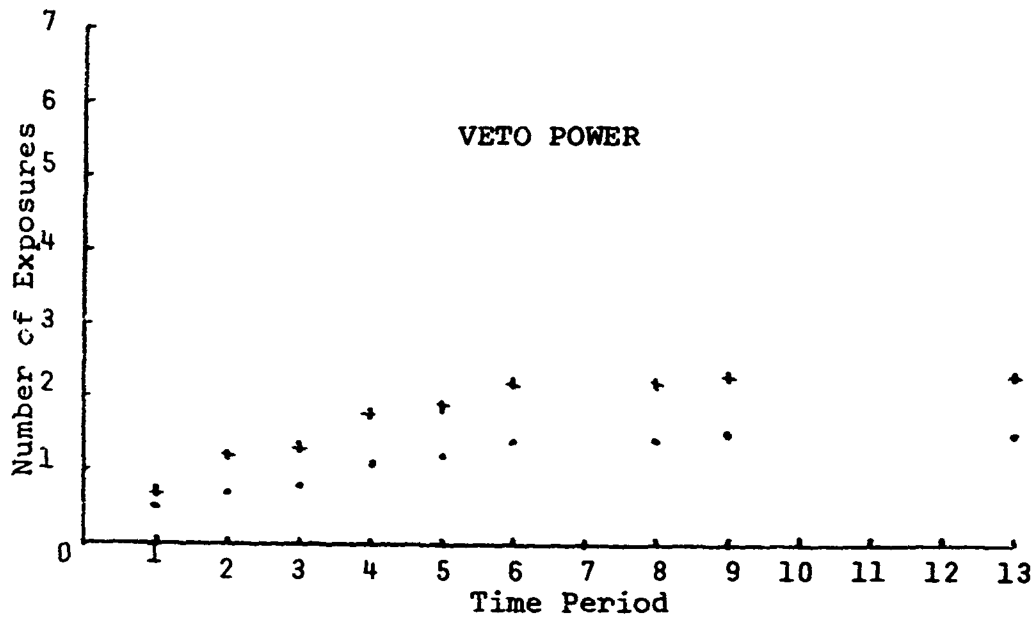


Figure VIII-20. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 5

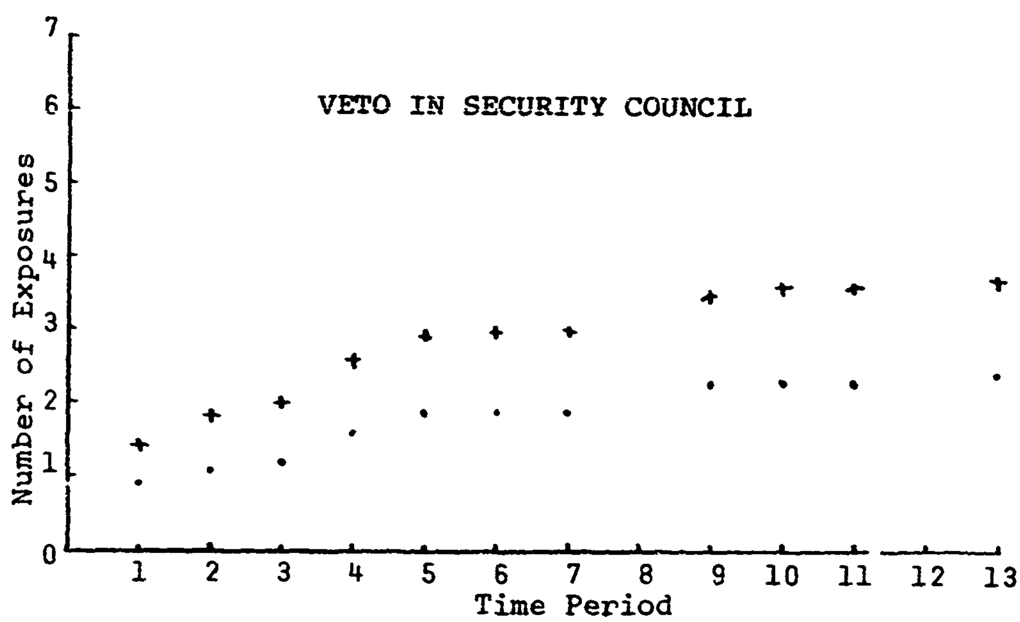


Figure VIII-21. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 6

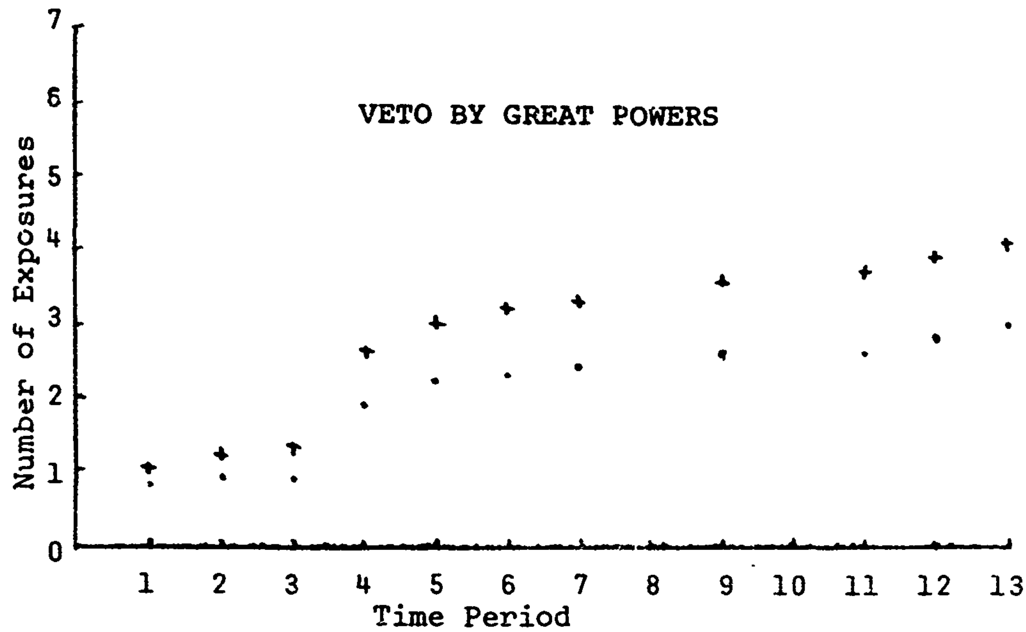


Figure VIII-22. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 7

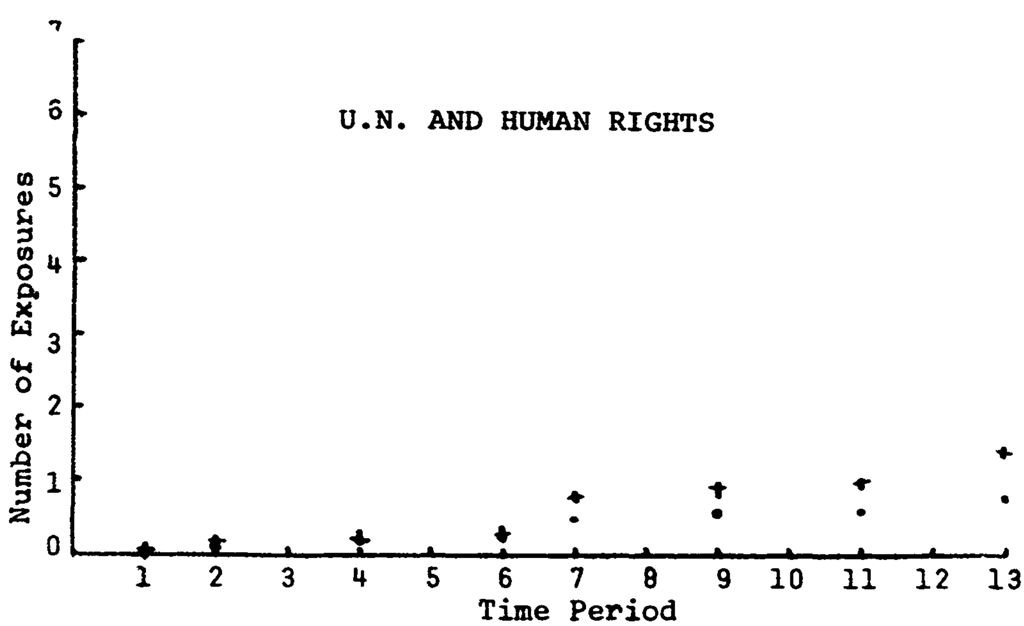


Figure VIII-23. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 8

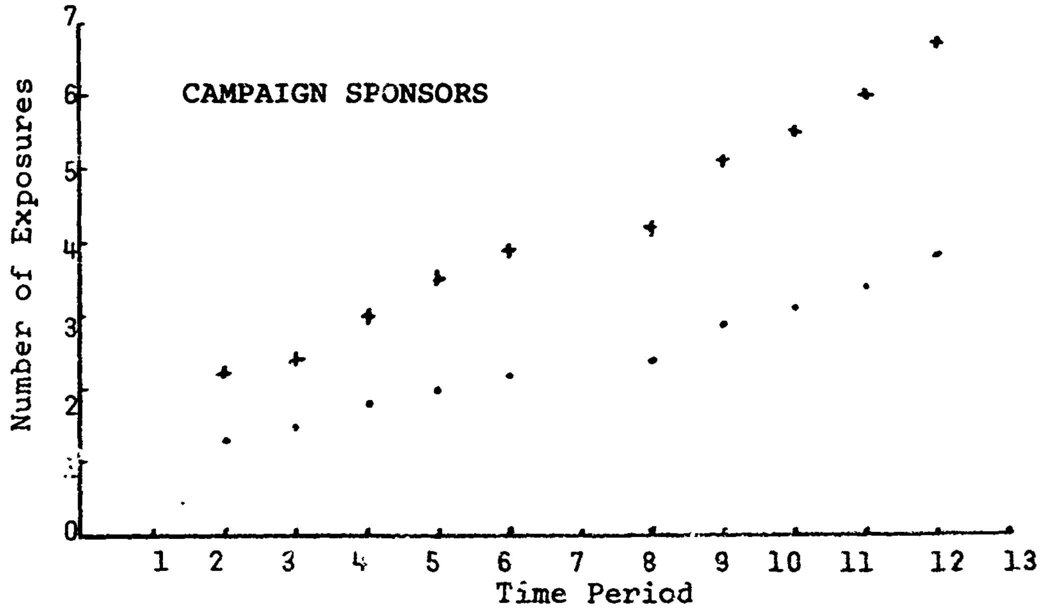


Figure VIII-24. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 9

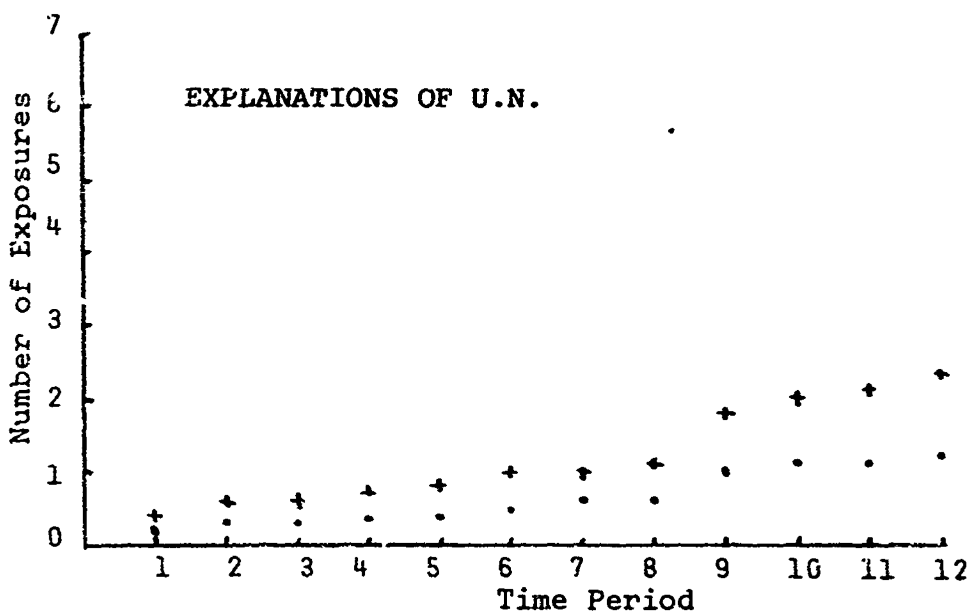


Figure VIII-25. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 10

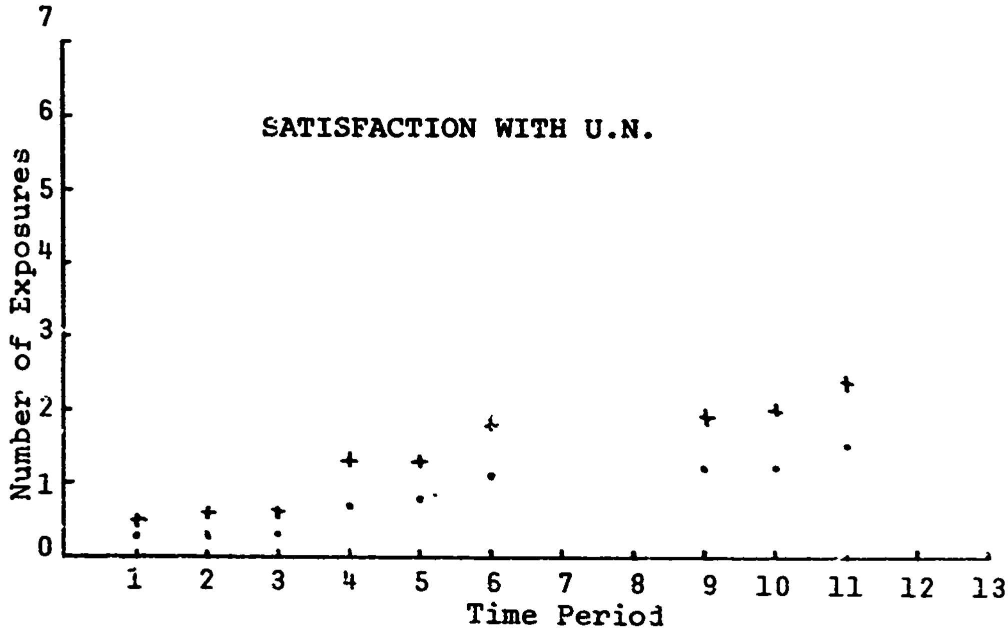


Figure VIII-26. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 11

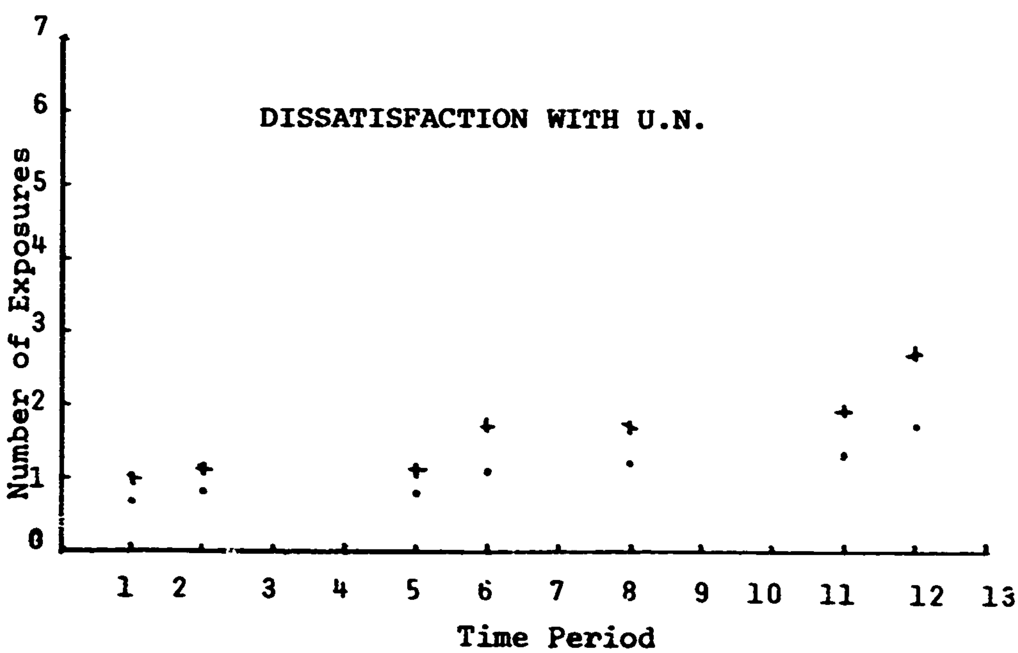


Figure VIII-27. The Average Cumulative Number of Exposures for Males (+) and Females (•) to Theme 12

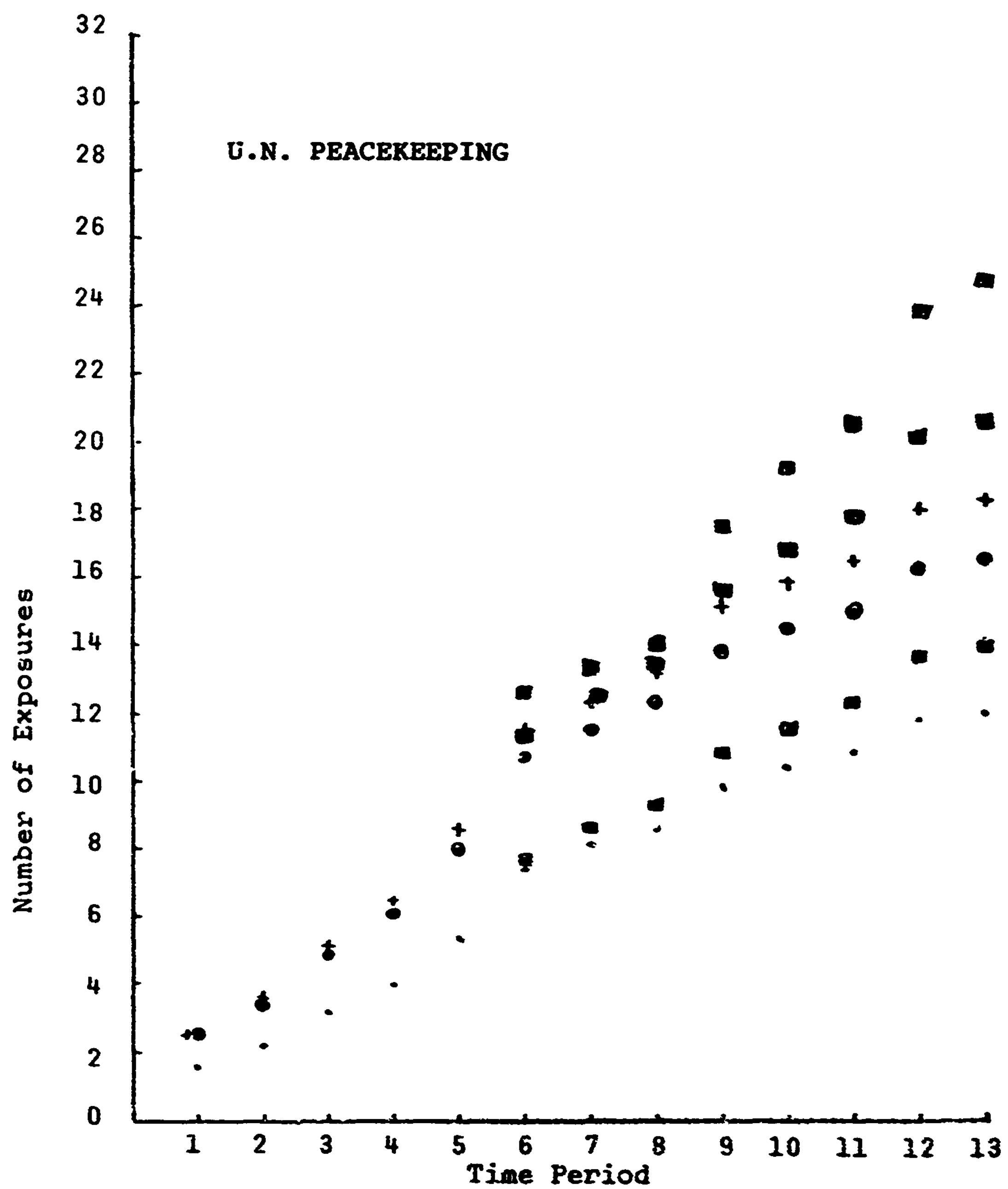


Figure VIII-28. The Average Cumulative Number of Exposures For College (+), High School (●), and Grade School (•) Education to Theme 1

(points in boxes are adjusted figures)

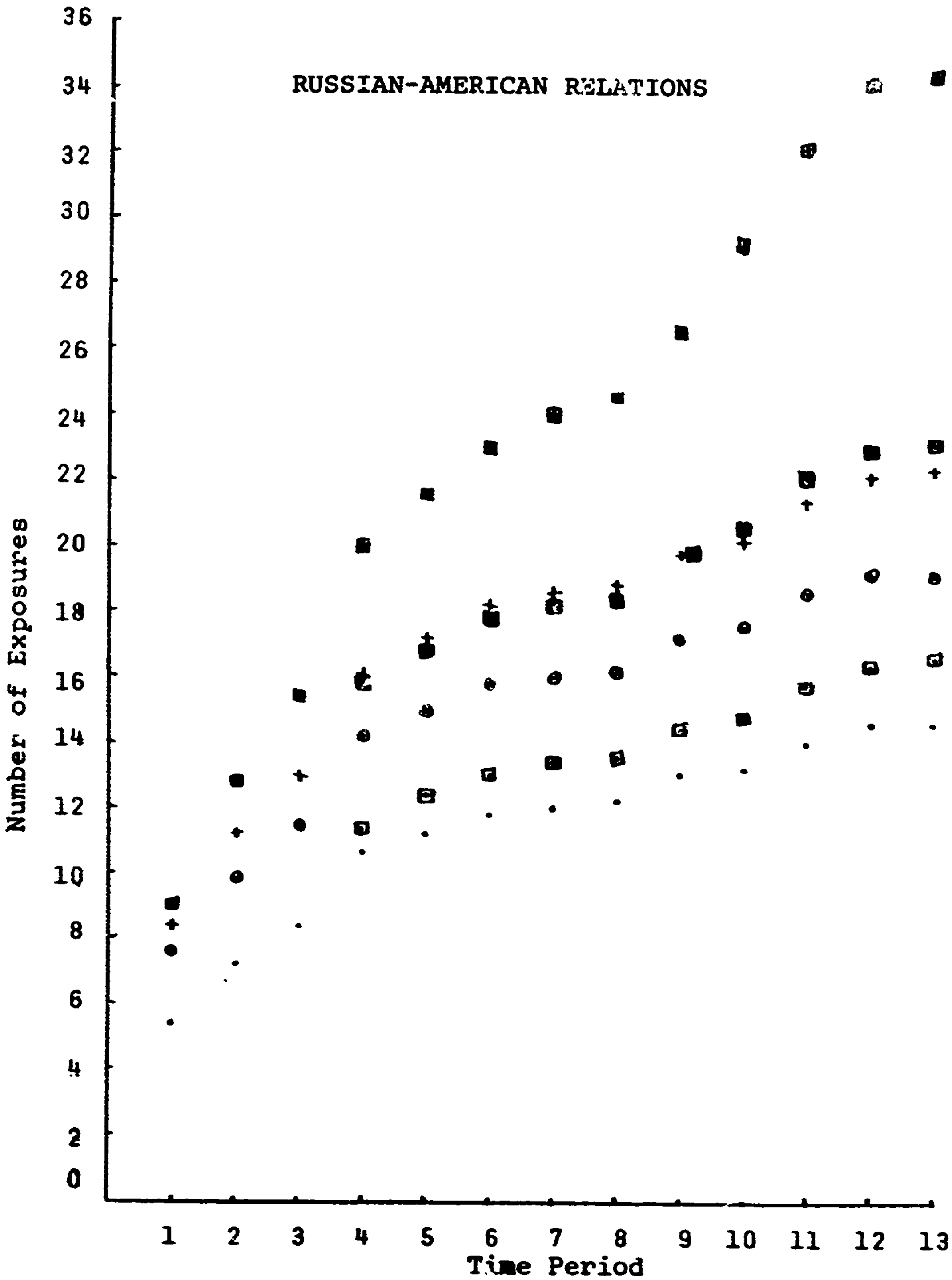


Figure VIII-29. The Average Cumulative Number of Exposures For College (+), High School (●), and Grade School (•) Education to Theme 2

(points in boxes are adjusted figures)

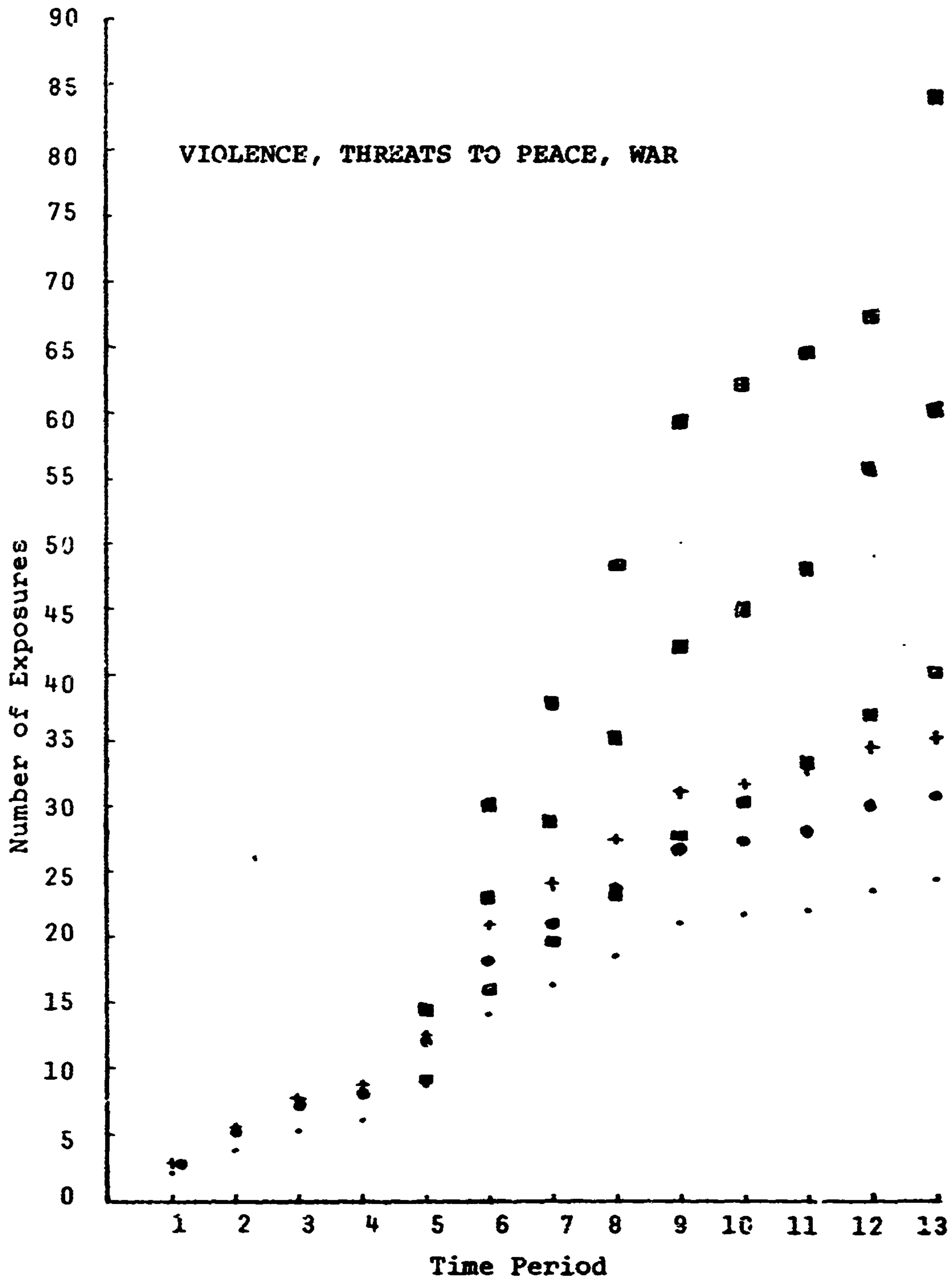


Figure VIII-30. The Average Cumulative Number of Exposures For College (+), High School (⊙), and Grade School (·) Education to Theme 3

(points in boxes are adjusted figures)

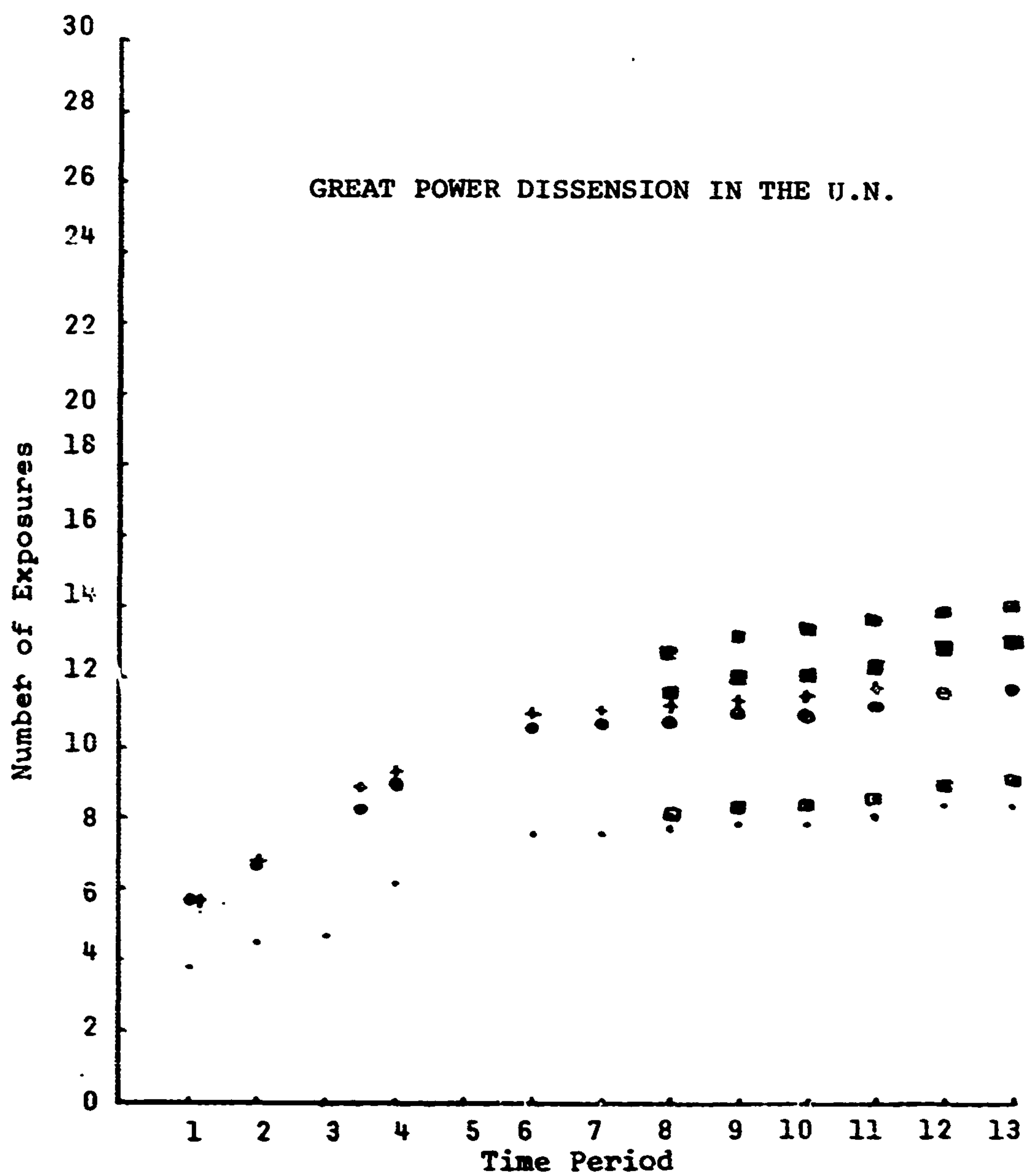


Figure VIII-31. The Average Cumulative Number of Exposures for College (+), High School (o), and Grade School (•) Education to Theme 4

(points in boxes are adjusted figures)

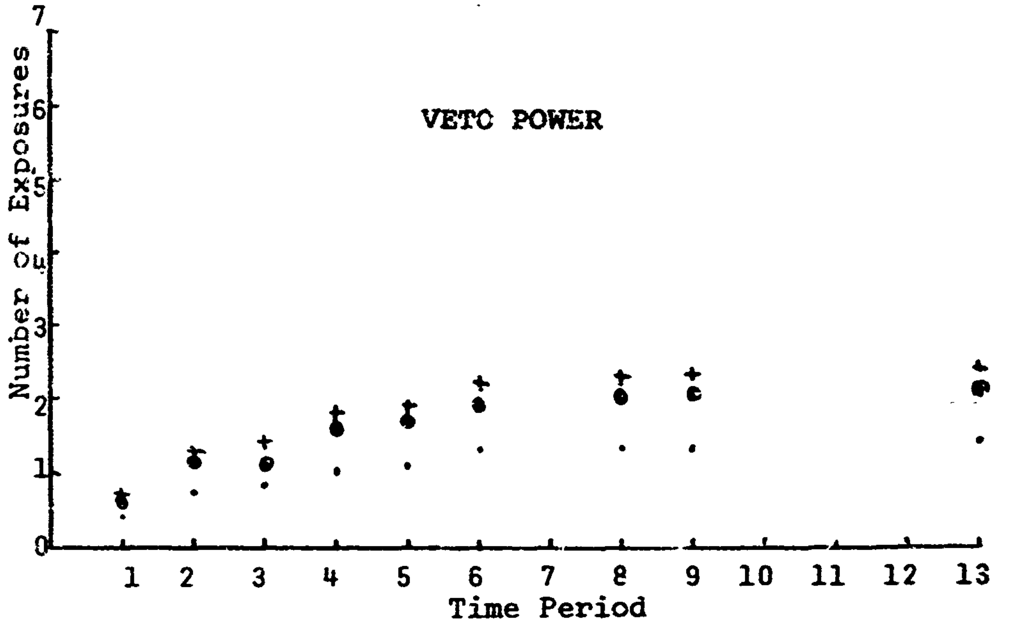


Figure VIII-32. The Average Cumulative Number of Exposures for College (+), High School (●), and Grade School (•) Education to Theme 5

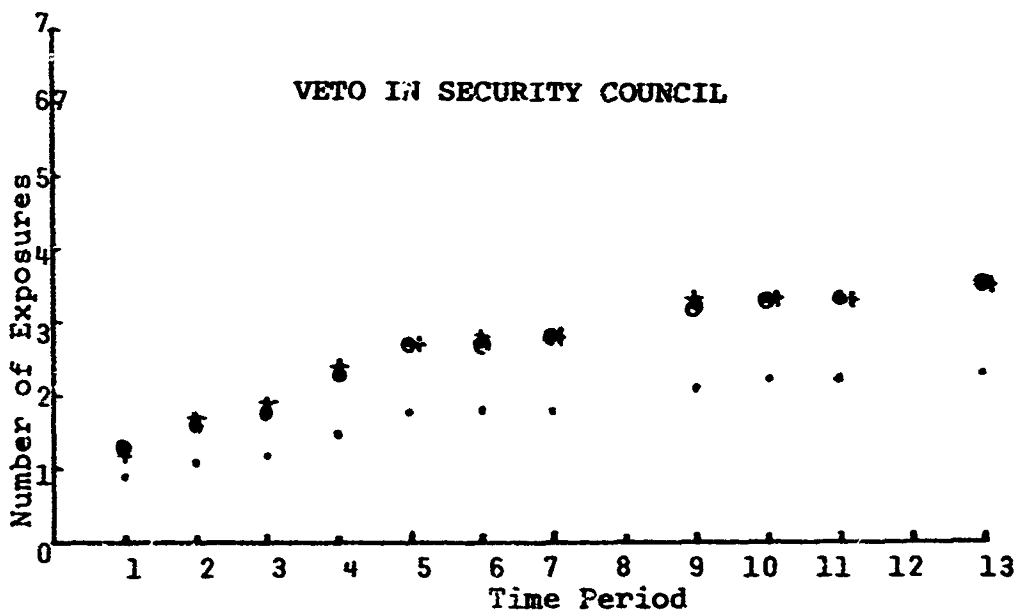


Figure VIII-33. The Average Cumulative Number of Exposures For College (+), High School (●), and Grade School (•) Education to Theme 6

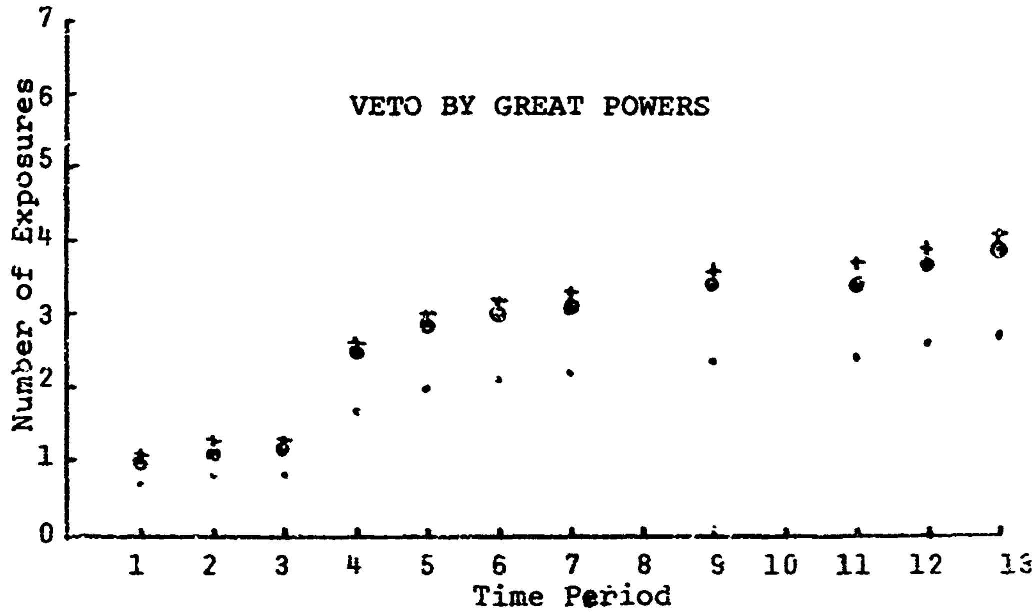


Figure VIII-34. The Average Cumulative Number of Exposures for College (+), High School (◉), and Grade School (•) Education to Theme 7

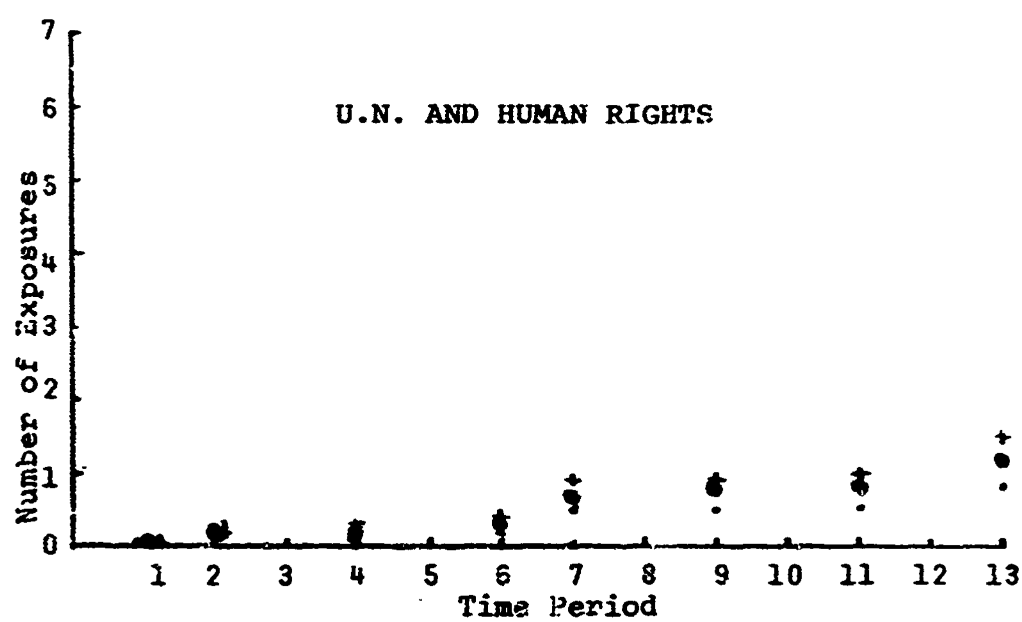


Figure VIII-35. The Average Cumulative Number of Exposures For College (+), High School (◉), and Grade School (•) Education to Theme 8

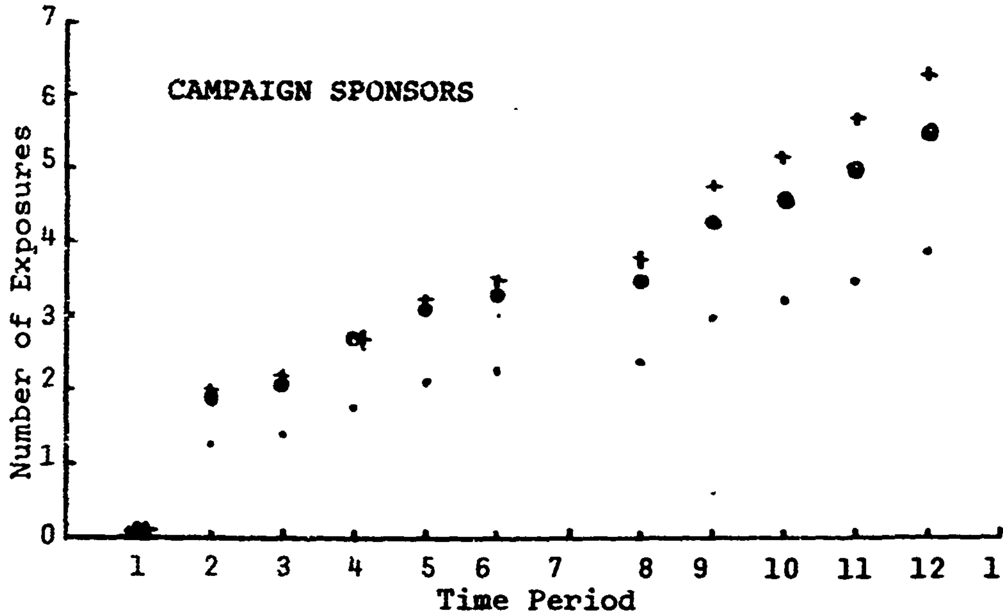


Figure VIII-36. The Average Cumulative Number of Exposures For College (+), High School (●), and Grade School (•) Education to Theme 9

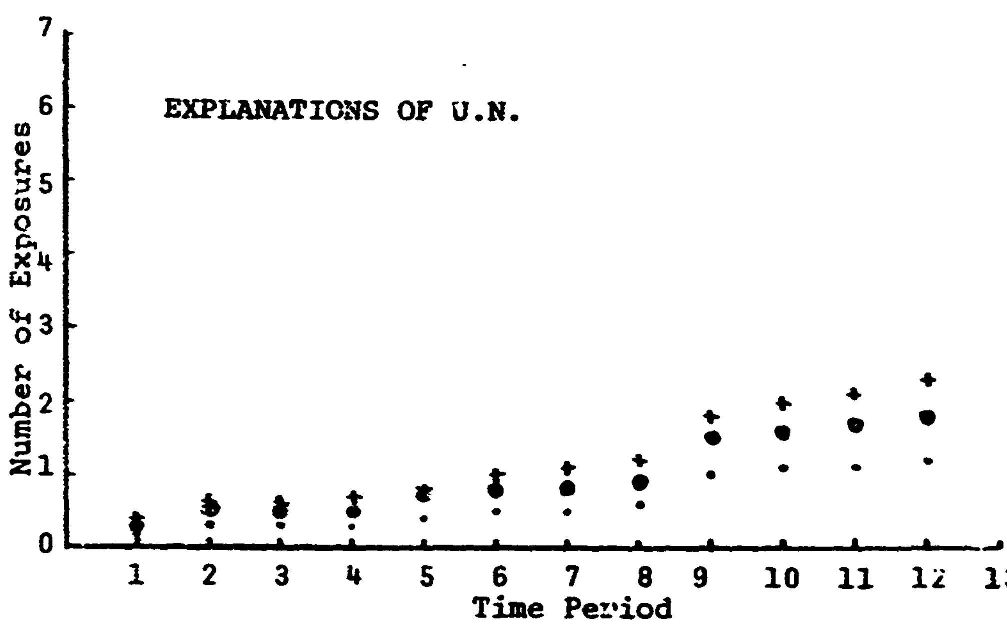


Figure VIII-37. The Average Cumulative Number of Exposures For College (+), High School (●), and Grade School (•) Education to Theme 10

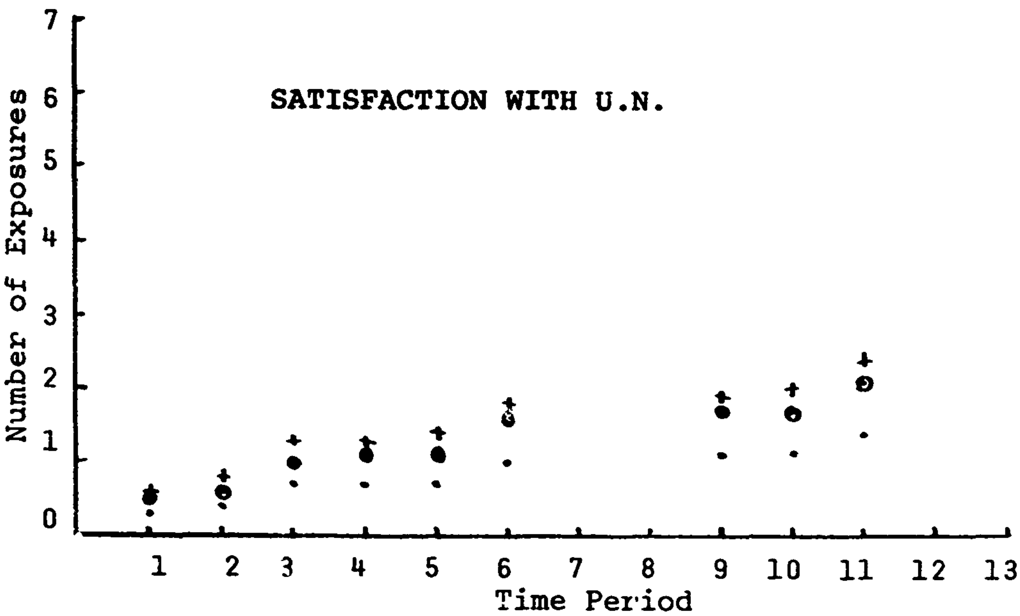


Figure VIII-38. The Average Cumulative Number of Exposures For College (+), High School (◎), and Grade School (•) Education to Theme 11

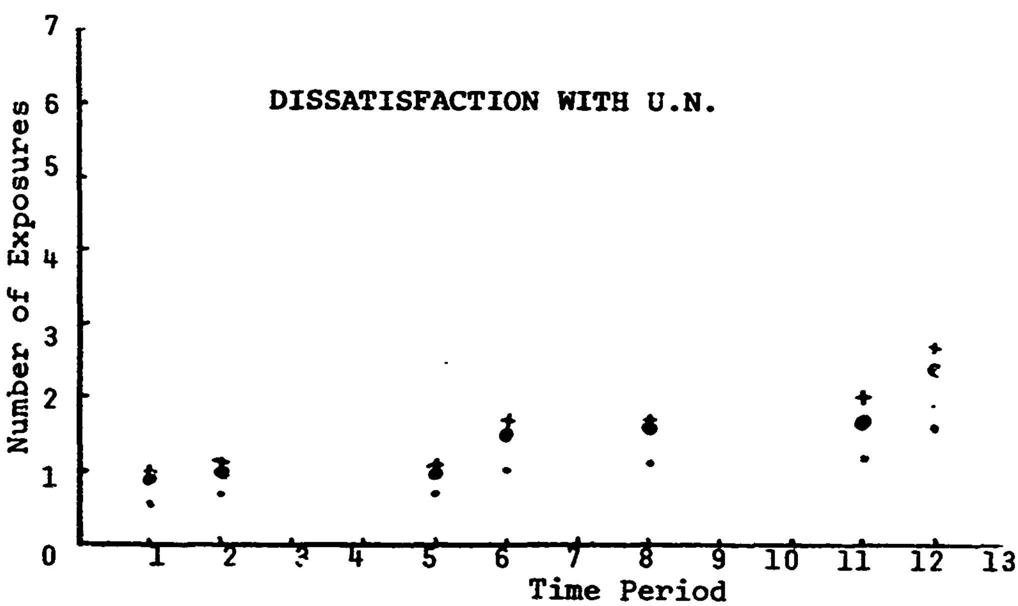


Figure VIII-39. The Average Cumulative Number of Exposures For College (+), High School (◎), and Grade School (•) Education to Theme 12

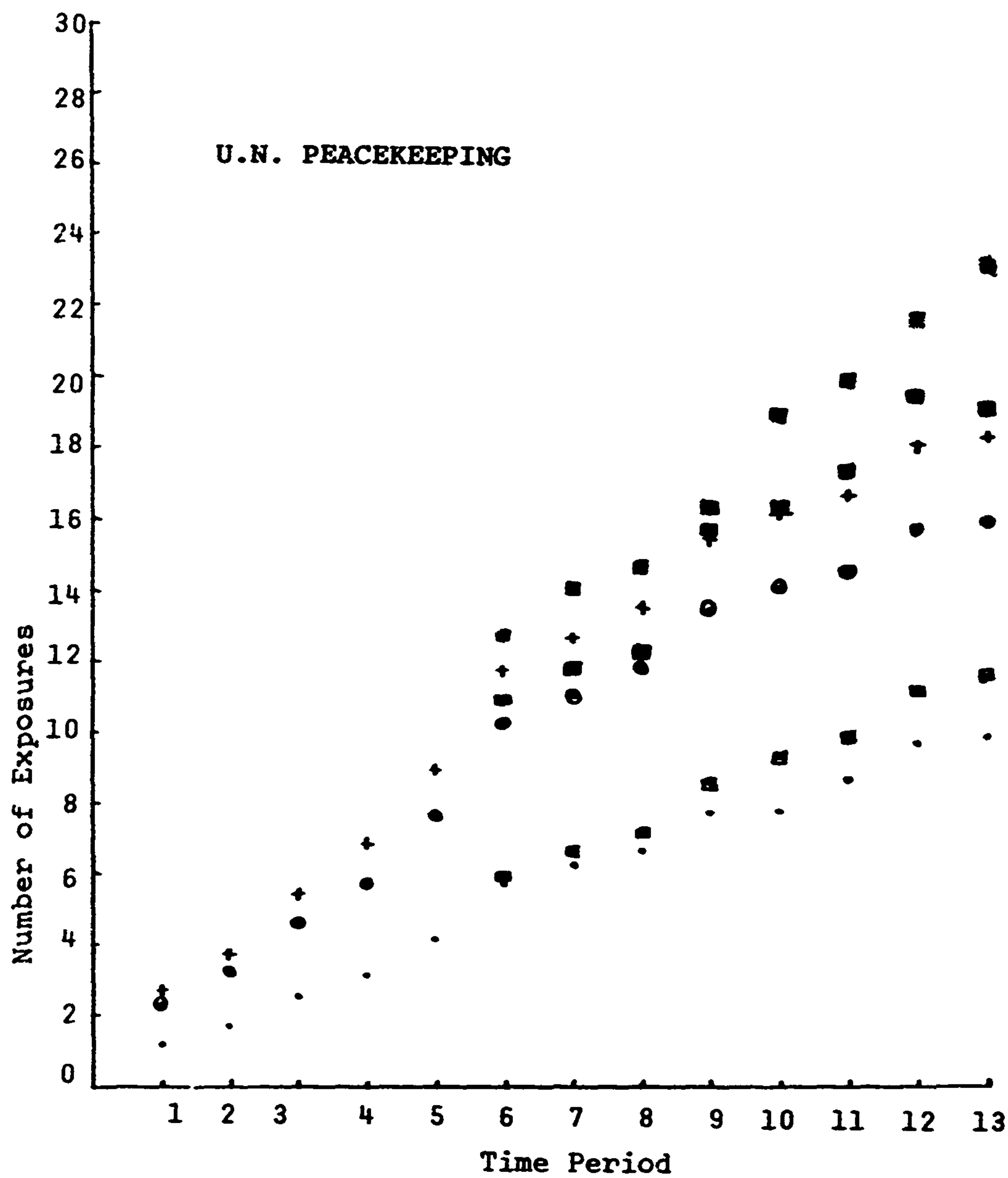


Figure VIII-40. The Average Cumulative Number of Exposures for High (+), Middle (⊙), and Low (·) SES to Theme 1

(points in boxes are adjusted figures)

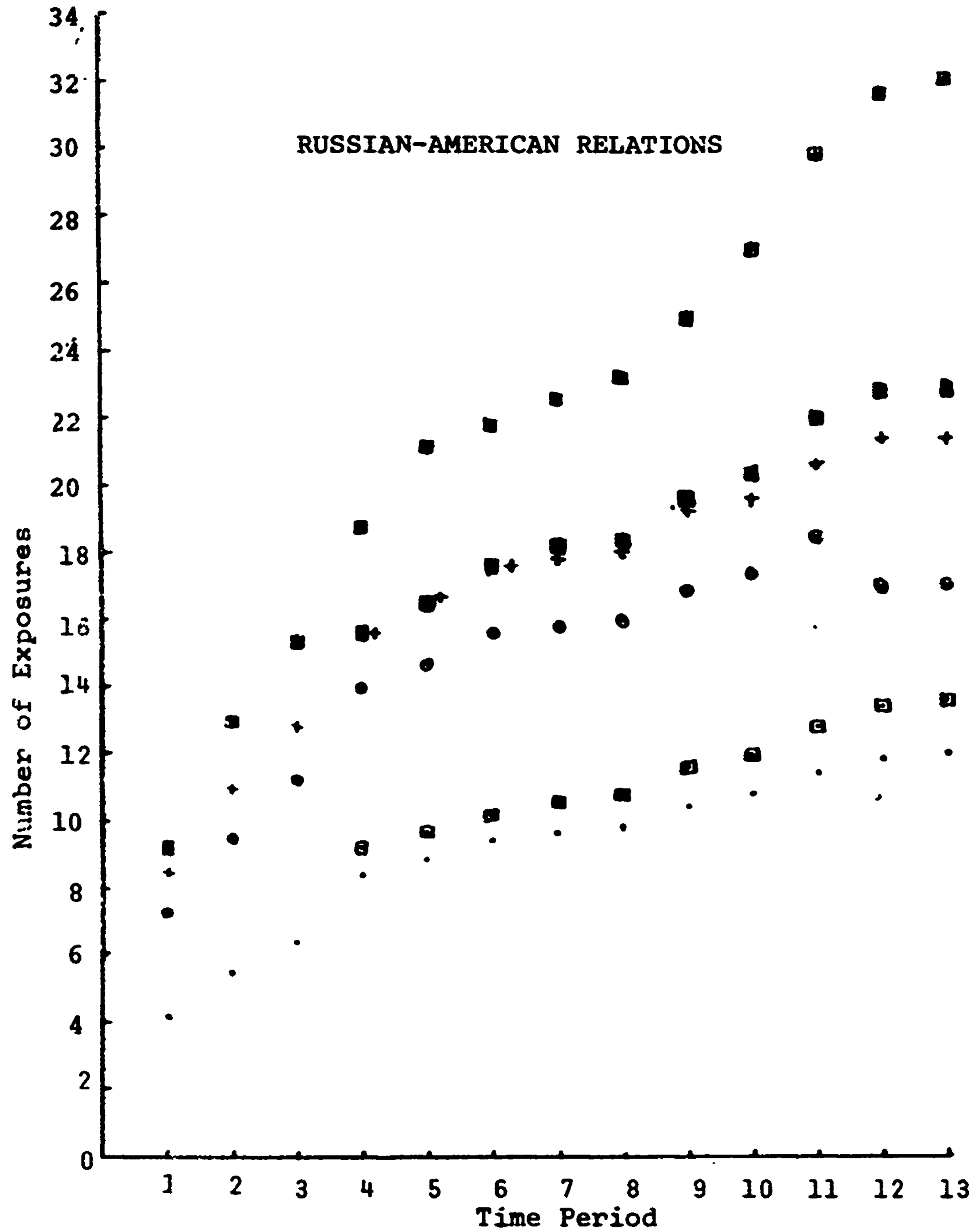


Figure VIII-41. The Average Cumulative Number of Exposures for High (+), Middle (⊙), and Low (•) SES to Theme 2

(points in boxes are adjusted figures)

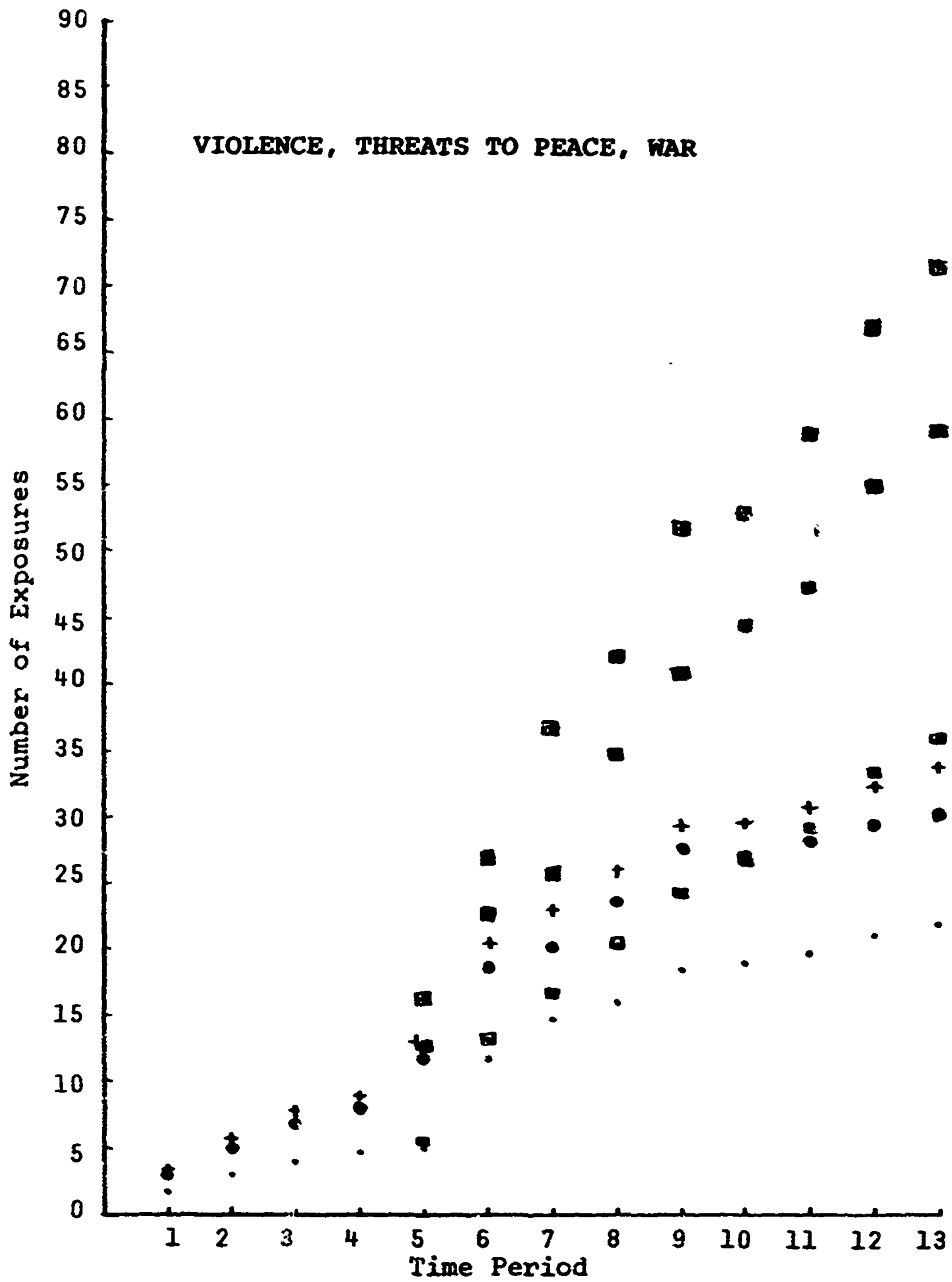


Figure VIII-42. The Average Cumulative Number of Exposures for High (+), Middle (⊙), and Low (•) SES to Theme 3

(points in boxes are adjusted)

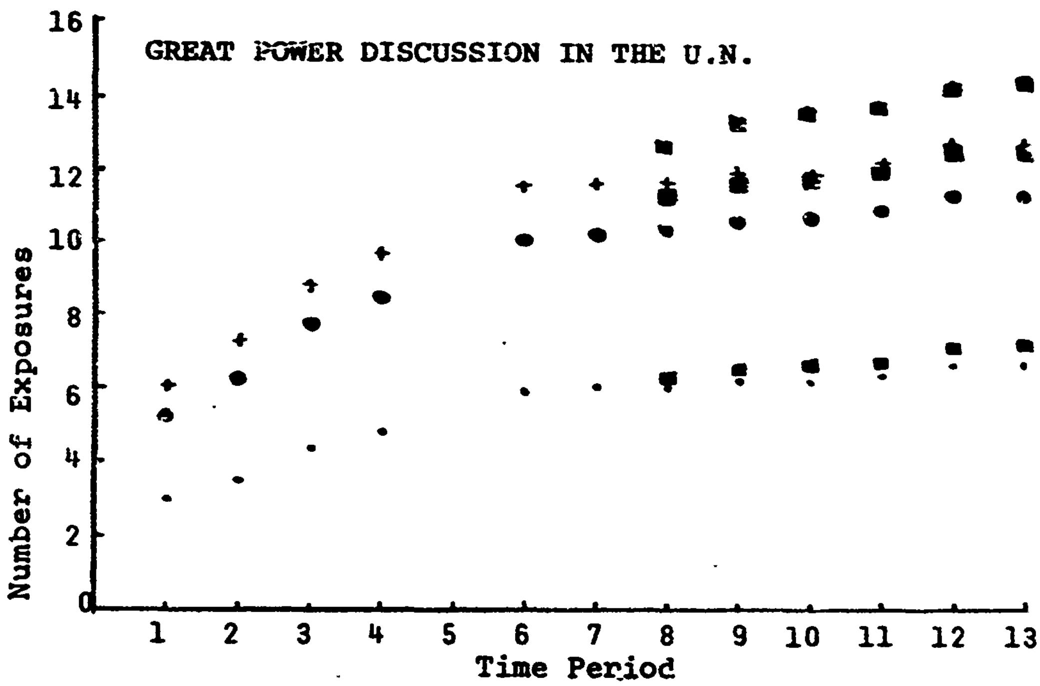


Figure VIII-43. The Average Cumulative Number of Exposures for High (+), Middle (●), and Low (·) SES to Theme 4

(points in boxes are adjusted figures)

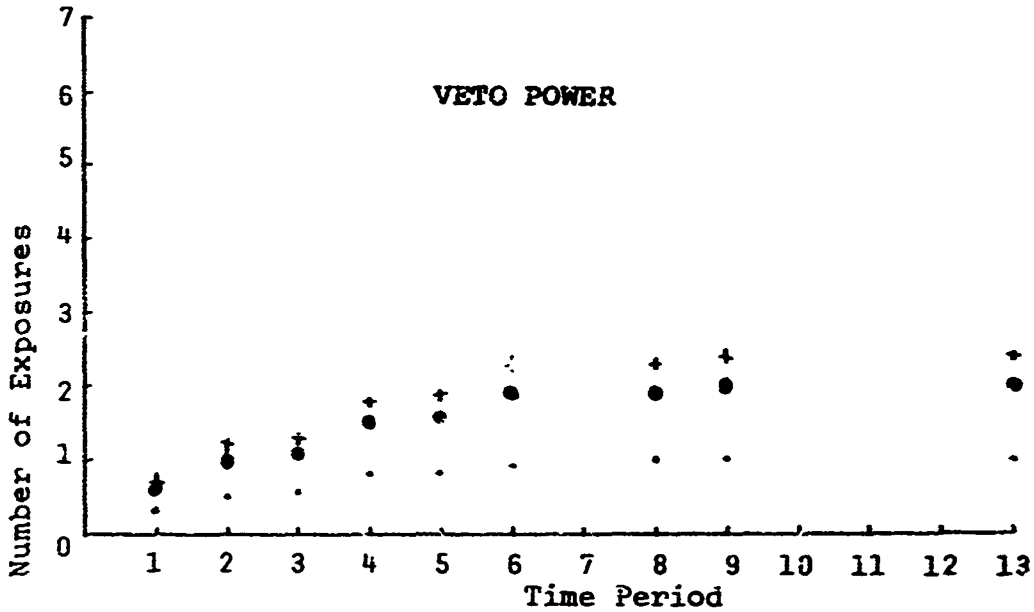


Figure VIII-44. The Average Cumulative Number of Exposures for High (+), Middle (●), and Low (•) SES to Theme 5

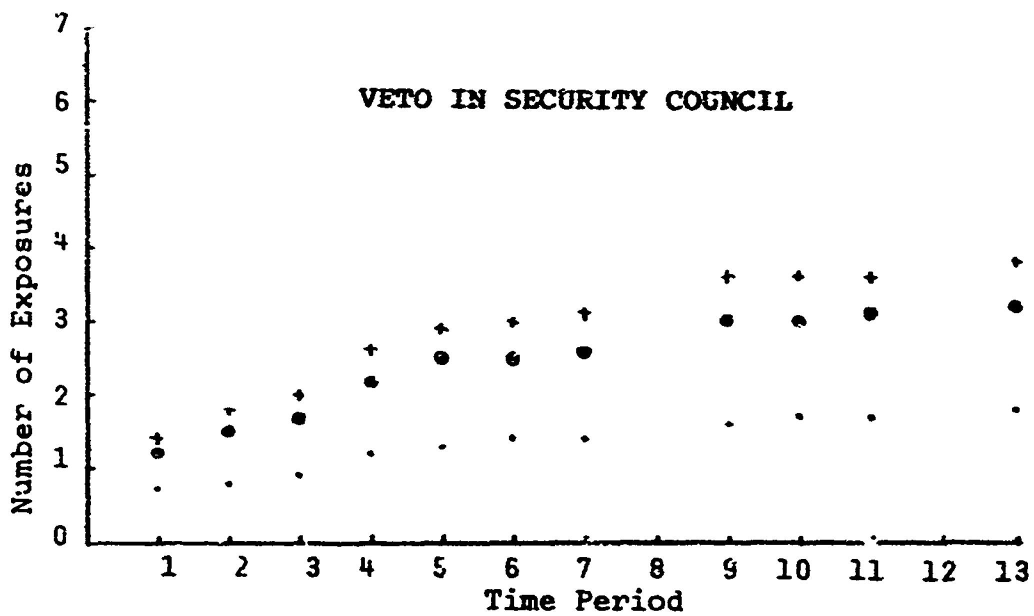


Figure VIII-45. The Average Cumulative Number of Exposures for High (+), Middle (●), and Low (•) SES to Theme 6

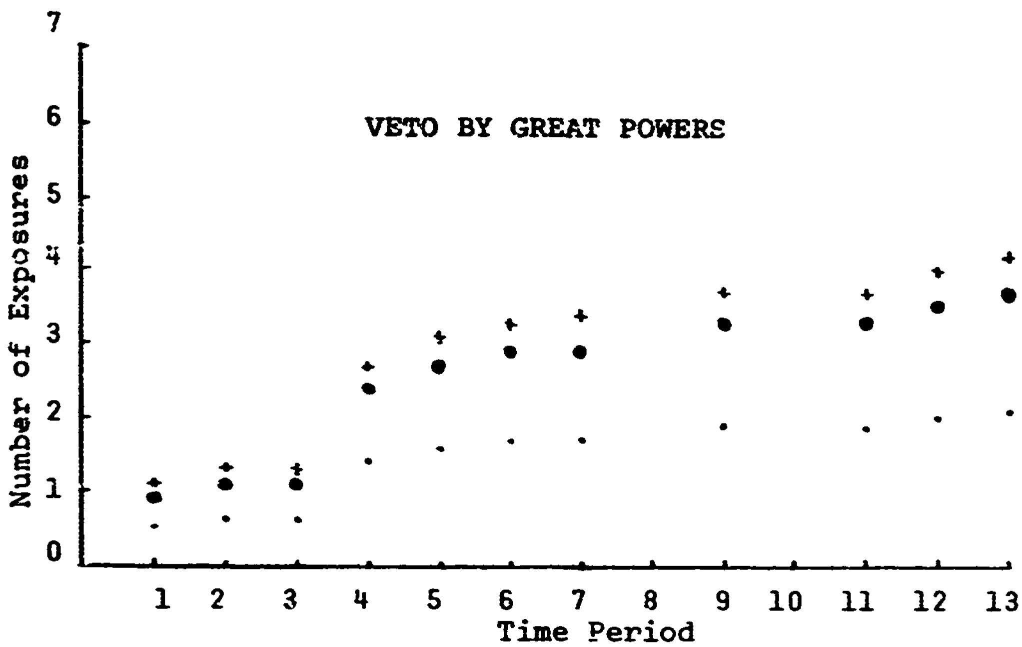


Figure VIII-46. The Average Cumulative Number of Exposures for High (+), Middle (●), and Low (•) SES to Theme 7

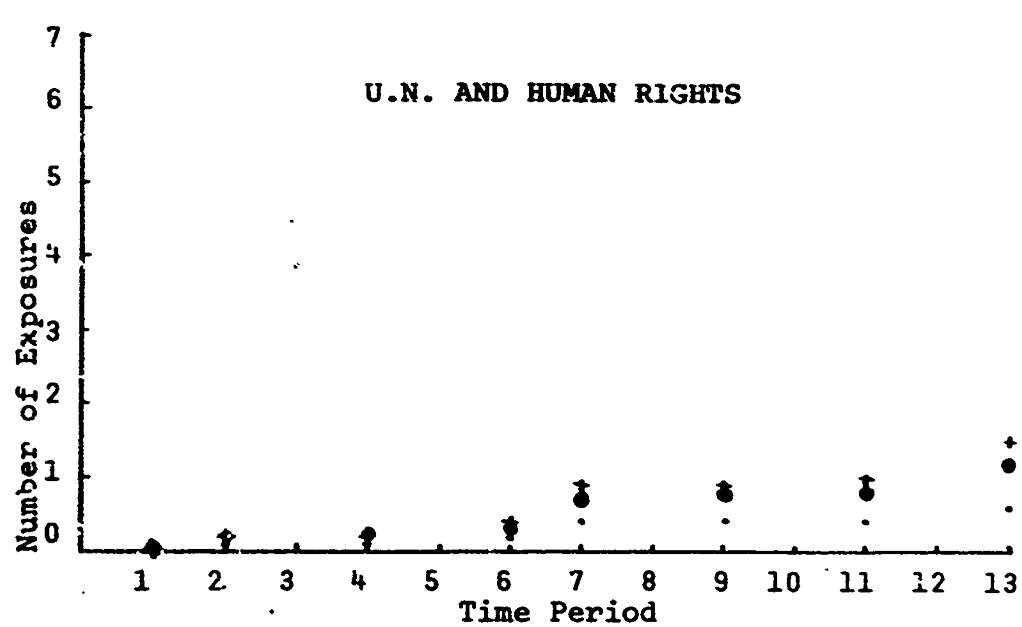


Figure VIII-47. The Average Cumulative Number of Exposures for High (+), Middle (●), and Low (•) SES to Theme 8

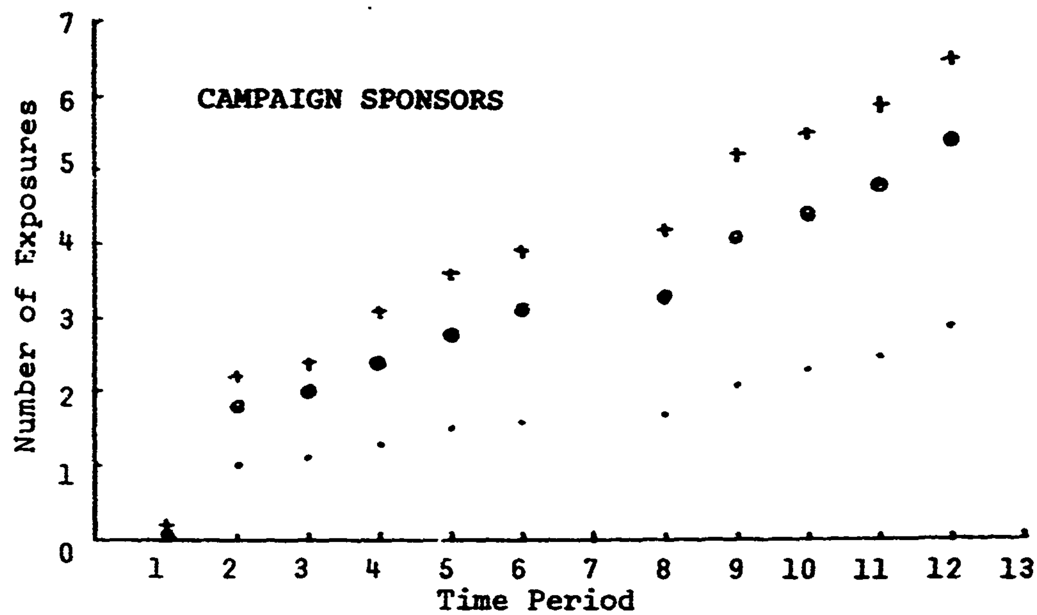


Figure VIII-48. The Average Cumulative Number of Exposures for High (+), Middle (●), and Low (•) SES to Theme 9

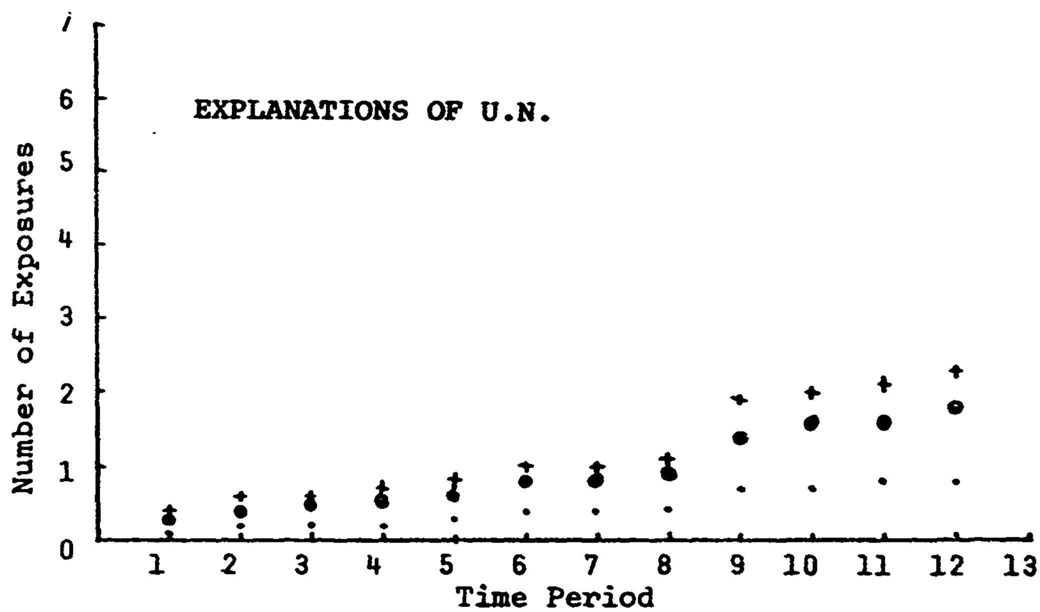


Figure VIII-49. The Average Cumulative Number of Exposures for High (+), Middle (●), and Low (•) SES to Theme 10

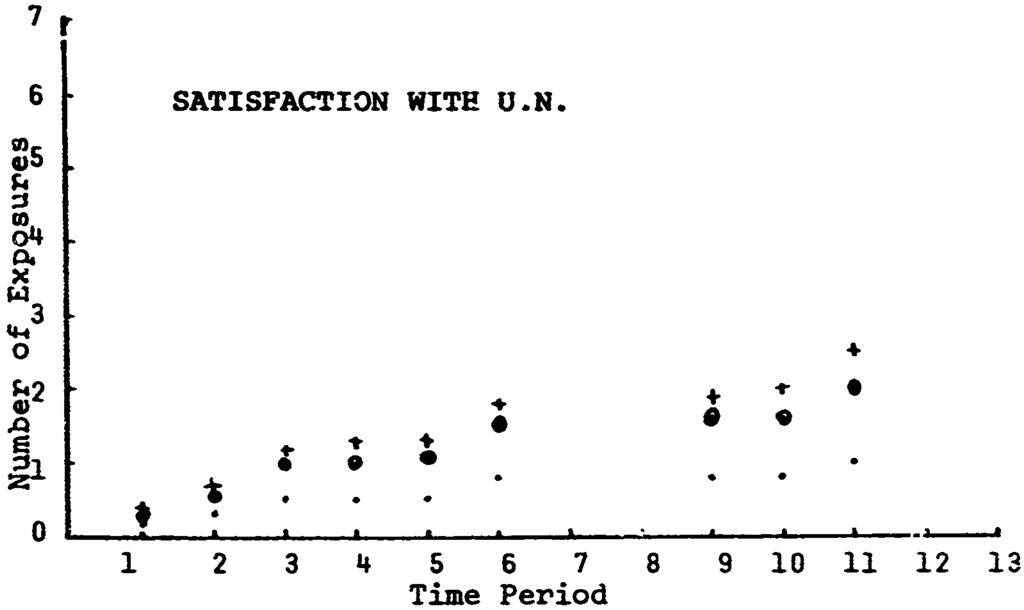


Figure VIII-50. The Average Cumulative Number of Exposures for High (+), Middle (⊙), and Low (•) SES to Theme 11

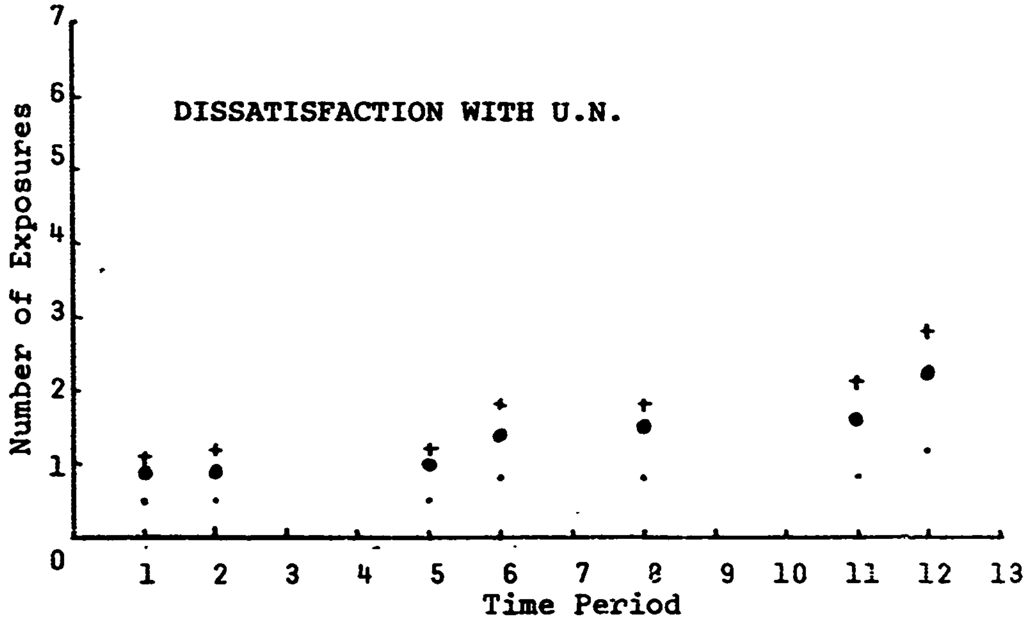


Figure VIII-51. The Average Cumulative Number of Exposures for High (+), Middle (⊙), and Low (•) SES to Theme 12

An Attempted Validation; Measuring the Changes in the Cincinnati Panel

In addition to the face validity of the internal logic of the model and the event validity of some of the assumptions of the model as tested against real cumulation data (this is described in the Appendix to Chapter V), and the plausibility of the records of exposures generated by the media system in both the trial and the real scenarios, we would also like to establish a more difficult validation; namely, the correlation of the expected number of exposures to various themes with the changes found in the NORC panel over the six-month period, across several population types. There are several difficulties in attempting this kind of correlation. First, we must define what we mean by change in the panel, somehow sorting out those changes which might reasonably be assumed to have resulted (in part) from exposure to messages in the media system from those changes which seem to be unrelated to exposure and are perhaps more or less random fluctuations in public opinion. Second, it is not at all obvious that one can construct some simple model of change as measured by survey data, that can be correlated with, or causally related to, exposure to messages in the mass media. Third, it is quite likely that exposure to the same messages in the mass media causes different directions of change in different audience types in the population, and indeed, in different individuals

within the same audience type.[11] We shall not attempt to account for changes in different directions, upon exposure to similar messages.

In addition, we must remember that the data from which the simulation model of the media system was constructed was itself imperfect--often inaccurate or lacking--and often had to be constructed from numerous assumptions and estimations. Likewise the content analysis was not as reliable as it might have been (recall that all the radio messages were inferred from the newspaper messages) and the generation of message exposure probabilities based on format factors, etc., at best explained only half of the variance in message exposure. It is also true (as we shall see below) that the themes coded in the content analysis are not always directly and obviously relevant to items in the NORC survey; occasionally the themes which are correlated are only marginally relevant in any obvious way to the opinions or information tapped by the questionnaire items.

Finally, we shall use the three dimensions, sex, educations, and socio-economic status, to define the audience types across which the correlations are calculated. (These

---

[11]Thus, although the dimensions by which the population is typed may explain a large proportion of the variance in audience <u>exposure</u>, they are not necessarily those dimensions which <u>well explain</u> <u>attitude</u> or <u>information change</u>.

dimensions were chosen because they produce the largest differences in exposure.) These three dimensions define eighteen audience types,[12] each of which we will divide into four cells on the basis of before-and-after responses to the questionnaire items, resulting in a total of seventy-two cells. Recall that the NORC survey panel consisted of approximately six-hundred respondents, which when spread across the seventy-two cells, produces an average of fewer than nine respondents per cell. Thus, all things considered, we have no great expectation of high correlations between the exposures predicted by the simulation and the changes found in the survey panel; however, insofar as the real media system is reasonably well modeled and errors in the construction of the media system plus effects unaccounted for are random, we should expect to find regular, small correlations between the exposure data and some measure of change in the panel.

With these caveats in mind, we proceed to the correlation of exposure to themes in the simulation with the changes in information or opinion measured in the NORC survey panel. We must decide which questionnaire items appear to be closely related to which themes in the

---

[12]Actually we shall use only 16 of the 18 cells in the correlations because two of the cells, males and females of college education and low SES, have only two and one members, respectively, in the simulation population.

scenario. We will attempt to show that changes in responses to the questionnaire items correlate positively with exposure to those themes where we perceive a relationship between the theme and questionnaire items, and that for other questionnaire items and other themes which are not related, the correlations are random. Thus, we test on the one hand for significantly large correlations where expected, and on the other hand for statistically insignificant correlations. The first step, however, is to construct some index of change of attitude or opinion in the population.

Several Indices of Opinion and Information Change

One can think of several possibilities for indices of change applied to panel data. For instance, let us suppose that the item asks some question about a matter of fact for which there is a definite correct answer. An index might measure the change within a group of people as the difference between the proportion initially giving the correct answer and the proportion finally giving the correct answer. Thus, if sixty per cent give the correct answer initially and seventy-five per cent give a correct answer finally, then there has been a net positive change of fifteen per cent. It is well known that this measure of change has a bias, however, i.e., for groups in the population which initially begin at a very high proportion of

correct answers, the possibility for positive change, in terms of proportion of the group which can make such a change, is limited by the initially high proportion of correct answers. This is the ceiling effect, which is commonly a problem in measuring such political phenomena as conversion or change of opinion in a political campaign.

One way to avoid this ceiling effect is to normalize the measure, i.e., divide the proportional increase by the maximum possible proportional increase from the first to the second time. If we let P_1 and P_2 represent the proportions giving the correct answer in September and in March, respectively, then this first index of change is just

$$\text{Index}_1 = \frac{P_2 - P_1}{1 - P_1}$$

We can calculate the value of Index_1 from the entries in a fourfold table describing the pattern of response to an item, initially and finally, in any group or audience type. For the table in Figure VIII-52, the initial proportion correct is $P_1 = (a + c)/N$ and the final proportion is $P_2 = (a + b)/N$.

		Initial Responses		
		Correct	Incorrect	
Final Responses	Correct	a	b	a + b
	Incorrect	c	d	c + d
		a + c	b + d	N

Figure VIII-52. A Pattern of Possible Responses to an Item at Two Times

Thus Index_1 is given by

$$\text{Index}_1 = \frac{\frac{a+b}{N} - \frac{a+c}{N}}{1 - \frac{a+c}{N}}$$

$$= \frac{\frac{b-c}{N}}{\frac{N-(a+c)}{N}}$$

or $$\text{Index}_1 = \frac{b-c}{b+d} \qquad \text{(VIII-1)}.$$

A second possible index of conversion or change is formed in the following way: we calculate the proportion of those giving the wrong answer in September who give the correct answer in March and subtract from this proportion the proportion who seem to be guessing. In terms of the fourfold table the proportion of those wrong in September who are correct in March is $b/(b + d)$, and the proportion guessing (the proportion of those who were

correct in September and wrong in March) is estimated as $c/(a + c)$. The second index is then

$$\text{Index}_2 = \frac{b}{b + d} - \frac{b}{b + d} \cdot \frac{c}{a + c}$$

or, simplified,

$$\text{Index}_2 = \frac{b}{b + d} \cdot \frac{a}{a + c} \qquad \text{(VIII-2)}.$$

A third index can be calculated, incorporating the notion of guessing, but extending it by recognizing that even among those people who gave correct answers in both September and March, a small proportion were probably guessing. Let w represent the true number who know the correct answer in September (and assume that they are recorded correctly in the survey), and let x represent the true number who do not know in September but who guess with an average probability p of guessing correctly. Also let y and z represent the true numbers in those two groups in March. If we make the assumption that no one who truly knows in September forgets or "unlearns" in the interval from September to March, then the true number who learn is just y-w and the proportion of those who did not know the correct answer in September, who have learned in the interval, is (y-w)/x. This is the third conversion or change index which must be expressed in terms of the observed quantities a, b, c, and d.

The observed number correct in September is just (a + c), which equals the number truly knowing in

September (w) plus the proportion who do not know but guess correctly $(p \cdot x)$. Thus

$$a + c = w + px \qquad \text{(VIII-3)},$$

and, using similar reasoning,

$$b + d = x - px \qquad \text{(VIII-4)},$$

$$a + b = y + pz \qquad \text{(VIII-5)},$$

and

$$c + d = z - pz \qquad \text{(VIII-6)}.$$

Also the number giving incorrect answers both in September and March (d) is just the proportion of those not knowing even in March, who guessed wrong on both occasions, $(1 - p)^2 z$,

$$d = (1 - p)^2 z \qquad \text{(VIII-7)}.$$

These five equations involving the five unknowns w, x, y, z, and p, must now be used to express the index $(y-w)/x$ in items of a, b, c, and d.

First we form $(y-w)$ by subtracting equation (VIII-3) from equation (VIII-5) and rearranging terms

$$y - w = b - c + p(x-z) \qquad \text{(VIII-8)}.$$

Subtracting equation (VIII-6) from equation (VIII-4) we find

$$b - c = x - z - p(x-z)$$

or

$$x - z = \frac{b - c}{1 - p} \qquad \text{(VIII-9)}.$$

Substituting equation (VIII-9) in equation (VIII-8) gives

$$y - w = b - c + p\frac{(b - c)}{(1 - p)}$$

$$= \frac{b - c}{1 - p} \qquad \text{(VIII-10)}.$$

From equation (VIII-7)

$$z = \frac{d}{(1 - p)^2}$$

and substituting for z in equation (VIII-6)

$$c + d = (1 - p)z$$

$$= \frac{d}{1 - p}$$

$$1 - p = \frac{d}{c + d} \qquad \text{(VIII-11)}$$

and

$$p = \frac{c}{c + d} \qquad \text{(VIII-12)}.$$

Substituting 1 - p of equation (VIII-11) in equation (VIII-10) we find that the real number who have learned from September to March is

$$x - z = \frac{(b - c)\ (c + d)}{d}$$

$$= (b - c)\ (1 + \frac{c}{d}) \qquad \text{(VIII-13)}.$$

Thus the best estimate of the number learning from September to March is slightly larger than the first obvious estimate, (b - c). We must next solve the equations for x, the number not knowing in September. Rearranging equation (VIII-4)

$$x(1 - p) = b + d$$

or $$x = \frac{b + d}{1 - p}$$

and substituting equation (VIII-11) for the term (1 - p),

$$x = \frac{(b + d)\ (c + d)}{d} \qquad \text{(VIII-14)}.$$

Thus the third change index becomes (from equations (VIII-13 and VIII-14))

$$\text{Index}_3 = \frac{y - w}{x}$$

$$= \frac{(b - c)\ (c + d)}{d} \ \frac{d}{(b + d)\ (c + d)}$$

or $$\text{Index}_3 = \frac{b - c}{b + d}$$

Therefore, both Index_1 from the first simple model, and Index_3 derived from the more sophisticated model, have the same values for any fourfold table. Table VIII-17 shows values of the three indices for several distributions of responses. For example, the first row shows the case of maximum change, i.e., all the initial responses were incorrect and all the final responses were correct. For this case, a, c, and d are zero; Index_1 and Index_3 equal 1.0 but Index_2 is undefined although it could logically be set equal to 1.0 at this limit. In general, the three indices do not differ much, especially if the ratio c/b is close to 1.0. Because of the appeal of the model which underlies Index_3, we have used this index for measuring change in the survey panel. The next task, then,

is to express the survey data in the form of fourfold tables to which the index can be applied.

Table VIII-17. Several Values of Three Indices of Change in Fourfold Tables for Certain Special Cases

Amount of Change	Values of Indices: Index_1, Index_3 $\left(\frac{b-c}{b+d}\right)$	Values of Indices: Index_2 $\left(\frac{b}{b+d}\right)$ $\left(\frac{a}{a+c}\right)$
Maximum change a, c, d, = 0	1.0	undefined
One-half of those initially wrong change; no guessing c = 0; b = d	.5	.5
One-half of those initially wrong change, but one half guess c = a, b = d	$\frac{1}{2} \cdot \left(1.0 - \frac{c}{b}\right)$	.25
All those initially wrong change, but all guess a = 0, d = 0	$1.0 - \frac{c}{b}$	0.0
No change at all c = 0, b = 0	0.0	0.0

The Preparation of the Survey Data: The Problem of the "Correct" answer.

The index of change described above assumes that the data can be put in the form of a fourfold table in which one of the two possible answers to the item is a correct answer, in the sense that it is an answer toward which information or attitudes are changing. If the question

is indeed a factual question, and only the correct answer may be presumed to be abroad in the mass media, then the model applies relatively directly. For example, Table VIII-18 below shows the September and March responses to a factual question asking whether it was the job of the United Nations to protect Human Rights. After eliminating respondents for whom one or both of the answers was not ascertainable, the answers to this factual question fall into the three categories, "Yes," "No," and "Don't Know." Table VIII-18 shows that there was a slight increase (from 79.5 to 83.3 percent) for the correct answer ("Yes") and corresponding decreases in the "No" and "Don't Know" responses. In this case, we have created a fourfold Table VIII-19 below from the ninefold table by combining into one category the "No" and "Don't Know" responses. Thus in all the items relating to factual matters, the "Don't Know" responses, if any, were combined with the incorrect responses to reduce the responses to a dichotomy.

The situation is less clear when we turn to matters of attitude or opinion. Do the "Don't Know" responses actually reflect the respondent's lack of information or interest or do they represent a position on a continuum somewhere between a "Yes" response and a "No" response, i.e., inability of the respondent to decide? In these cases we resolve the matter in the following manner: In

Table VIII-18. A Ninefold Table of Responses Concerning the United Nations and Human Rights

March Responses	September Responses Yes	No	D.K.	Marginals[a]
Yes	85.8%	67.8%	92.9%	83.3%
No	10.7	28.2	7.1	13.4
D.K.	3.4	7.1	0.0	3.4
	100.0%	100.0%	100.0%	
	(1977)	(397)	(112)	
Marginals[a]	79.5%	16.0%	4.5%	100.0% (N=2486)

[a]The row and column marginals are based on the total number of weighted responses (N=2486). However the total number of respondents in the panel was 592 of which thirty percent were filtered out prior to this question because they could not identify the United Nations.

Table VIII-19. A Fourfold Table of Responses Concerning the United Nations and Human Rights

March Responses	September Responses Yes	No-D.K.	Marginals[a]
Yes	85.8%	73.3%	83.3%
No-D.K.	14.2	26.7	16.7
	100.0%	100.0%	
	(1977)	(509)	
Marginals[a]	79.5%	20.5%	100.0% (N=2486)

[a]The row and column marginals are based on the total number of weighted responses (N=2486).

the ninefold table we compare the marginal responses for September and March to determine which of the responses, "Yes" or "No," has increased in frequency during the interval and call this response the "correct" response. Thus we assume that public opinion encouraged by the media output, is shifting in the "correct" direction. Next we note the direction of change of the "Don't Know's." If the "Don't Know's" also increase during the interval then they are combined with the correct response; if they decrease from September to March then they are combined with the incorrect response.

As an example, suppose that the percentage of people expecting war increased from 30 to 60 percent during the six months. If the percentage of "Don't Know's" also increased (say) from 10 to 15 percent, while the percentage not expecting war decreased from 60 to 25 percent, we make the most plausible assumption that the "Don't Know's" represent people responding to the same forces which seem to be increasing the expectation of war. Thus, they are combined with the category expecting war.

This solution increases the amount of change in the table over that which one would find if one simply eliminated all the "Don't Know" responses. Of course, some of the questionnaire items seemed to have responses which fall naturally into a dichotomy, e.g., a question was asked about several subjects to determine whether or not the respondent had a "keen" interest in the subject.

In addition to those items for which an index of change has been calculated there were also several items or indices which could be directly correlated with certain themes in the simulation. For example, one of the indices formed was an index of change in information over the six-month period from several factual items relating to international affairs. This index can be directly correlated with exposure to themes in the scenario. Also, several of the March questions asked about the respondent's exposure to news about the United Nations during the previous six months; for these questions we have coded the number and kinds of vehicles mentioned by the respondents as exposure vehicles in the previous six months. These items are then directly correlated with exposure to the scenario themes. In Table VIII-20 we have listed the variables measured by the NORC survey which have been used in the correlations with exposure and the net changes in these variables.

How is Exposure to Themes Related to Conversion or Change?

What index of exposure shall we attempt to correlate with recall of exposure or with the index of change in the panel? Given that we are running the correlations over sixteen population subgroups, an obvious candidate for the index of exposure is the average expected number of exposures for the population subgroup. This first

Table VIII-20. NORC Panel Measures Showing Attitude and Information Change and Mass Media Exposure over the Six Months

Abbreviation	Item Paraphrase	Code for Responses		Proportion Giving First Response[a] March	Proportion Giving First Response[a] September
General Political Issues and Opinions					
PROB-WAR	Problems facing U.S. now?	Fighting another war, keeping peace	Other	.450	.237
EXPECT WAR	Expect the U.S. to fight another war within the next ten years?	Yes	No-D.K.	.722	.494
PROB-USSR	Problems facing U.S. now?	Relations with Russia, Russian foreign policy	Other	.363	.172
COUNT ON USSR	Can we count on Russia to meet us half-way in working out problems?	No	Yes-D.K.	.777	.759
INTEREST-USER	How much interest in news about relations with Russia?	Keen	Other	.660	.544

Table VIII-20. (Continued)

Abbreviation	Item Paraphrase	Code for Responses		Proportion Giving First Response[a]	
				March	September
INTEREST-BOMB	How much interest in news about the control of the atomic bomb?	Keen	Other	.565	.516
HIGH INFO CHANGE	Index of improvement in answering five factual data questions	+3 to +5	All Other	.120[a]	
MEDIUM INFO CHANGE	Index of improvement in answering five factual data questions	+1 or +2	All other	.264	
<u>Measures of Exposure to the U.N. Over the Last Six Months</u>					
MEDIA-NEWSPAPERS	During the last six months have you seen anything in the newspapers about the U.N.?	Yes	No, or not asked because respondent could not identify main purpose of U.N.	.816	
MEDIA-RADIO NEWS	During the last six months have you heard any radio newsprograms about the U.N.?	Yes	No, or not asked	.629	

Table VIII-20. (Continued)

Abbreviation	Item Paraphrase	Code for Responses		Proportion Giving First Response[a]	
				March	September
MEDIA-SIGNS	During the last six months have you seen any signs or posters about the U.N.?	Yes	No, or not asked	.242	
MEDIA-LEAFLETS	During the last six months have you read any leaflets or pamphlets on the U.N.?	Yes	No, or not asked	.133	
MEDIA-MEETINGS	During the last six months have you been to any meetings where the U.N. was discussed?	Yes	No, or not asked	.155	
MEDIA-OTHER	During the last six months have you seen or hear anything else about the U.N.?	Yes	No, or not asked	.227	
MEDIA-NUMBER	Number of media exposed to in previous questions (several other questions have been omitted).	Three to eight media	Less than three media or not asked	.377	

Table VIII-20. (Continued)

Abbreviation	Item Paraphrase	Code for Responses		Proportion Giving First Response[a] March	Proportion Giving First Response[a] September
Knowledge of Various Aspects of the U.N.					
KNOW U.N.	What is main purpose of the United Nations organization?	Keep peace, promote harmony, discourage aggression	D.K.	.754	.718
U.N. JOB-RIGHTS	Is it U.N. job to guarantee all people equal rights?	Yes	No-D.K.	.833	.795
U.N.-JOB-HEALTH	Is it U.N. job to improve world health conditions?	Yes	No-D.K.	.796	.755
U.N.-JOB-TRADE	Is it U.N. job to improve trade between countries?	Yes	No-D.K.	.733	.697
U.N.-JOB-BOMB	Is it U.N. job to deal with disarmamant and control of the atomic bomb?	No	Yes-D.K.	.412	.353

Table VIII-20. (Continued)

Abbreviation	Item Paraphrase	Code for Responses		Proportion Giving First Response[a] March	Proportion Giving First Response[a] September
UN VETO	Have you heard anything about the veto power in the United Nations?	Yes	No-D.K.	.446	.356
Interest in and Attitudes Toward the U.N.					
INTEREST-U.N.	How much interest in news about the United Nations?	Keen	Other	.332	.313
U.N. PROGRESS	In general, are you satisfied or dissatisfied with the progress of the U.N. thus far?	Dissatisfied	Satisfied D.K.	.555	.528
U.N. PROGNOSIS	Do you think U.N. will succeed in spite of disagreements, or will fail?	Other	Succeed	.489	.312
U.S. U.N. POLICY	Should U.S. work through U.N. or outside of U.N. to preserve peace?	Outside-D.K.	Through U.N.	.211	.185

Table VIII-20. (Continued)

Abbreviation	Item Paraphrase	Code for Responses		Proportion Giving First Responsea March	Proportion Giving First Responsea September
U.N. INTEREST NOW	Compared with six months ago, how much interest do you have in the U.N.?	None	The same-Less	.331	
MORE-PUBLICITY	Why do you think you become more interested?	Because U.N. is receiving more publicity	Other	.040	
MORE-CRISIS	Why do you think you have become more interested?	Because the situation has become more serious, critical	Other	.209	
KNOW GROUPS	Do you know names of any groups here in Cincinnati that help the U.N.	Yes	No-D.K.	.126	.067
GROUPS HELP	Do these groups help the U.N. in any practical way?	Yes	No-D.K.	.683	.578

Table VIII-20. (Continued)

Abbreviation	Item Paraphrase	Code for Responses		Proportion Giving First Response[a]	
				March	September
U.N. SLOGAN	Have you heard the slogan: 'Peace be-begins with the United Nations: the United Nations begins with you'?	Yes	No, or not asked	.463	

[a]Several items below have only one entry in (between) these columns. These items either measure net change over the six months, or recall of exposure during the time period.

model postulates that the amount of change shown in a subgroup is a linear function of the average number of exposures to a relevant theme by members of that subgroup. This is the first model actually tested with the data; however, this model ignores the likely presence in the subgroup of people who initially know the correct response. These people, according to present communications research, are likely to be more exposed than those initially not knowing and therefore the average number of exposures for the informed and the uninformed should differ. A second model posits a linear relationship between the average number of exposures of those initially uninformed in the subgroup and the index of change for the subgroup.

We have experimented with this second model as well as with the first, estimating the number uninformed in a subgroup as equal to the number initially giving the wrong response, and assuming that the exposure rate of the informed is 1.5 or 2.5 times that of the uninformed.

Finally an even more realistic model of change would posit some threshold of exposure necessary for a person to learn and report learning to an interviewer. With this assumption, we explore below four levels of complexity of this model.

Models of Conversion or Change Within Groups

The model of change suggested above involves two conceptual elements which together combine to describe the number of people changing as a function of the average number of exposures. The first of these elements is the range of thresholds of conversion present in the group of people initially not knowing. This range comes about either because at the onset of the simulation the hypothetical people are assumed to have different initial exposure levels, and/or because of intrinsic differences in thresholds due to the present attitude structure, cognitions, etc. In general we will find a distribution of those not knowing along a range of thresholds. This distribution might be a constant density of individuals, or it might bell-shaped, or might assume any one of a number of different shapes. The second element is the net exposure probability of the individuals to the messages in the theme, which governs their individual rates of growth of expected exposures. Let us explore how these elements combine to imply several different kinds of relationships between different changes or conversion indices and average number of exposures. We explore below several cases using hypothetical values and relationships between the two elements.

Case One. Suppose that each individual has the same threshold number of exposures for a change to occur, and

in addition, that all persons have identical net exposure probabilities for each message in the theme. If this is true, then the expected number of exposures for each individual, which is just the sum of the net exposure probabilities for the individual, all grow at exactly the same rate and the average expected number of exposures for all those not knowing is just the actual value of the expected number of exposures for each individual in that group. If this is the case, then all the expected number of exposures grow together toward the threshold and all of them pass the threshold at the same time; we find no change at all prior to that message which makes each person's expected number of exposures greater than or equal to the threshold, and one hundred percent change at that particular message. Thus we would expect, in this case, to fine that the change index has only two possible values, either zero or one (eliminating the possibility of guessing, of course), and the relationship between average exposure and change is a step function. The correlation between the change index and the average number of exposures would be quite small if the average number of exposures is spread out over a significant range for the cells over which we are correlating.

Case Two. In this case let us assume that there is a range of thresholds, that the individuals not knowing are distributed uniformly over this range, but as in the

previous case the net exposure probabilities are all equal over all the hypothetical persons not knowing. In this case, all the expected number of exposures grow together toward the lowest value of the threshold range but no change occurs until the person with the lowest threshold of change has an expected number of exposures equal to that threshold. Thus, the change index is zero up to this point. Once we have reached this point, the number of people changing will be proportional to the increment in expected number of exposures so long as the total expected number of exposures for the individual does not exceed the high end of the range. If it exceeds the range, then all the people have crossed the threshold and there is no change even though the number of exposures continues to grow. (There is also the possibility that the range is so small that the next message after one enters the range, takes one out of the range, or in fact that one jumps completely over the range, in which case we have a step function again.) Thus, in this case, we expect to have a linear increase in the change index with a linear increase in expected number of exposures, from the time one enters the threshold until the time one passes out of the threshold. Thus the correlation between the average number of exposures and the change index should be 1.0 within the range of thresholds but for exposures distributed much more widely than the range of thresholds

the correlation should be smaller, and the more widely the exposures are distributed the smaller the correlation.

<u>Case Three</u>. Another possibility is that the thresholds are all the same for all individuals, as in the first case, but that the net message exposure probabilities differ from individual to individual, from zero to one, and that in this range the not-knowing group are uniformly distributed. (This range of net exposure probability need not be from zero to one in this example, but this simplifies matters.) In this case, as a succession of messages is processed in the simulation, the individuals with the highest net message exposure probabilities grow quickly to the threshold and pass it, and the people with lower net message exposure probabilities grow more slowly. The net result is that as messages are processed there is a constant rate of conversion of individuals to the message, and a constant increase in the change index from zero to one with a constant increase in expected number of exposures, until the last person is converted.

<u>Case Four</u>. In the most complicated case, the uninformed members of the group are not uniformly distributed over the threshold range, but distributed in a normal curve or some other distribution. The net message exposure probabilities for the members are not equal, but differ from individual to individual, and as such we have

a distribution of net exposure probabilities. In this case as messages are processed, there are starts and jerks as people cross the threshold level of exposure. The resulting growth of converted people is certainly not linear with growth in average exposure, and in this most realistic case, the maximum correlation would certainly not be one.

Since we have no data on exposure thresholds for the simulated population and since the simulation does not store individual exposure rates, but only subgroup averages, we have not attempted to apply the four more elaborate models outlined above. The most realistic case does however, suggest that the relationship between average exposure and conversion in a group can be far from linear; thus we have additional grounds for expecting smaller correlations. In the next pages we look at the correlations resulting from the models directly relating average exposures for the subgroup or an estimate of average exposures for the uninformed in the subgroup to conversion or change.

The Effect of the Exposure Model On the Correlations Between Exposures and Changes in the Panel

In order to examine the consequences of the model transforming average raw exposures for a subgroup, we have collected together in Table VIII-21 several correlations of changes in the panel over the six months or

Table VIII-21. Correlations of Average Exposures to Explicit Explanation of the U.N. (Theme 10) with Several Opinions and Information Changes in the Cincinnati Panel over the Six Month Time Period

Simulation Time Period	High Information Change	Medium Information Change	Newspaper Exposure to the U.N.	Radio Exposure to the U.N.	Increased Knowledge of the Main Function of the U.N.	Increased Acquaintanceship with the Veto in the U.N.	Decreased Satisfaction with the Progress of the U.N.	Increased Proportion Expressing "Keen Interest in the U.N."	"More Interest in the U.N."	Increased Expectation of War Within Ten Years
Correlations with raw average exposure										
1	-.694	.655	.427	-.285	-.307	.598	-.565	.370	-.044	.125
3	-.707	.667	.413	-.336	-.344	.591	-.570	.362	-.095	.132
5	-.712	.686	.415	-.349	-.360	.592	-.551	.349	-.102	.145
7	-.718	.700	.419	-.356	-.374	.585	-.520	.329	-.107	.173
9	-.708	.669	.411	-.341	-.349	.582	-.567	[illegible]	-.114	.146
11	-.709	.667	.413	-.340	-.348	.586	-.576	.353	-.113	.138
13	-.710	.669	.417	-.336	-.348	.588	-.576	.349	-.111	.136
Correlations with individual[b] transformed average exposure (K=1.5)[a]										
13					-.364	.574	-.574	.382		-.000
Correlations with individual[b] transformed average exposure (K=2.5)[a]										
13					-.379	.551	-.561	.401		-.126
Correlations with "average"[b] transformed average exposure (K=1.5)[a]										
1	-.672	.651	.453	-.250	-.266	.609	-.643	.358	-.032	.030
3	-.685	.662	.441	-.296	-.301	.603	-.647	.351	-.078	.044
5	-.691	.683	.445	-.308	-.316	.606	-.631	.338	-.083	.057
7	-.700	.700	.452	-.315	-.332	.603	-.603	.320	-.088	.083
9	-.685	.664	.439	-.301	-.306	.595	-.644	.342	-.095	.058
11	-.686	.662	.440	-.300	-.305	.597	-.650	.342	-.094	.050
13	-.686	.662	.444	-.296	-.305	.599	-.651	.338	-.092	.049
Correlations with "average"[b] transformed average exposure (K=2.5)[a]										
1	-.639	.631	.465	-.216	-.224	.610	-.717	.344	-.025	-.055
3	-.651	.641	.456	-.255	-.254	.605	-.719	.337	-.064	-.047
5	-.658	.661	.462	-.265	-.269	.610	-.705	.325	-.068	-.037
7	-.668	.681	.472	-.271	-.283	.611	-.685	.309	-.072	-.014
9	-.651	.643	.455	-.259	-.259	.599	-.717	.329	-.079	-.036
11	-.652	.639	.455	-.258	-.259	.600	-.721	.329	-.079	-.041
13	-.652	.640	.457	-.255	-.258	.601	-.721	.325	-.077	-.043

[a] K is the ratio, used in the transformation, of the average number of expected exposures of knowledgeable persons to that of those less knowledgeable.

[b] Individual transformed average exposures are created for each theme, for each of the panel-measured variables for which a conversion index is calculated. "Average" transformed average exposures are created for each theme for the entire set of these panel-measured variables, based on the average number of knowledgeable and less knowledgeable persons in each population subgroup over the entire set of these variables.

recollections of exposure by the panel for the six months with the tenth theme of the scenario, explicit explanation of the United Nations. We chose this theme because the message content seemed particularly clear and definite and we might expect high correlations with changes in attitudes or information about the U.N., and also because the moderate number of messages in this theme and the resulting moderate overall average exposure produced negligible errors in recording average exposures. The error for this theme was only 5.1 exposures in a total of 3400 exposures.

First we look at the correlations of raw average exposure with several of the variables measured by the NORC survey.[13] These variables include the level of information change over the six month period (measured by the ability to answer five factual questions about the United Nations), high information change being a net improvement of from three to five factual answers and medium information change being net improvement of from one or two answers. Two more variables measure recall of exposure to information about the United Nations via the newspapers or radio. The next two variables are increased knowledge

[13]Although we have average exposures by each time period for a total of thirteen points at which the average exposure is measured, we have calculated the correlations only for every other time period beginning with the first, and ending with the thirteenth in order to simplify the presentation of the data.

of the main function of the United Nations, i.e., peace-keeping and acquaintanceship w:.h the veto power. The next variables measure decreasing satisfaction with the progress of the United Nations, the increasing proportion over the six months expressing keen interest in the United Nations, and the proportion at the end of the six months who say they have more interest in the United Nations than they did six months prior. The final variable is the increased expectation of war within the next ten years.

We noted in our discussion above of the model relating average exposure and conversion, that if we do not reach saturation as people are exposed to the messages of a theme, we might expect that the correlation between average exposures and the changes measured after six months would grow stronger as the time periods in the simulation progress. In other words, unless there is some sort of a saturation effect, we should expect that the relationship between the average exposures and changes found in the panel would be strongest after all the messages of the theme are run rather than at some intermediate point when only some part of the messages of the theme had been run. However, the data of Table VIII-21 indicate that there is some saturation or more complicated effect than the simple relationship between average exposure and change. None of the ten dependent variables

shows a consistent pattern of increasing correlations as the time periods progress; in general the correlations fluctuate somewhat from time period to time period about an average value. In fact, it does not seem that one time period offers a more respresentative correlation than any other time period and that we could have run the scenario through only one time period and gotten correlations with any one of the ten dependent variables of about the same magnitude and direction as we observe at the end of thirteen time periods. If the content analysis and the model of the mass media distribution system are valid, then we must conclude that there is a more complicated relation (e.g., a threshhold effect) between average exposure to a theme and opinion and information changes relevant to that theme over a six month period. The changes measured by the NORC survey cannot be linear functions of the average number of exposures for these subgroups.[14]

Another fact which probably diminishes the strength of the correlations is that none of the dependent variables are directly related to explicit explanation of the United Nations, the closest being the increased

---

[14]Another reason for the fact that the magnitude of the correlation does not grow as the time period increases, may be that there is a complicated process of assimilating and/or integrating the content of the messages over the six month period.

knowledge of the main function (peacekeeping) of the United Nations. The most direct measure with which we could correlate exposure to this theme would be a question asking people if they recall listening to a discussion or reading an article which helped to explain the purpose and functioning of the United Nations. Unfortunately we have no such question in the NORC data, and this recall would probably be quite unreliable even if we did. In addition, it may very well be that those people who would read or listen to messages offering an explicit explanation of the United Nations are those who already are quite familiar with it. If this is true we would not expect a high correlation between changes in knowledge and exposure to these messages.

On page 417 above we proposed a transformation of the raw average exposure to the theme in order to account for the fact that those people who initially had the prevailing knowledge or opinion would probably have a higher likelihood of being exposed to the theme while those with less knowledge would have a lower likelihood of exposure. Since those people with less initial knowledge account for the changes measured in the panel, we should correct the raw average exposure of the subgroup or population type to produce the average exposure only if those people who might be susceptible to change over the six month periods, i.e., those initially not

knowing the information or not having the opinion towards which other people were changing. These data transformations were performed as described above for each of the dependent variables for which there is a conversion index. The resulting averages, called individual transformed average exposures, have been calculated assuming that those already knowing have either one and a half times the average exposure of those not knowing (K = 1.5) or that they have two and a half times the average exposure of those not knowing (K = 2.5). The next two groups of correlations in Table VIII-21 show the resulting correlations with these individual transformed average exposures for each of the two values of K. We show these correlations for only five of the dependent variables because the other variables are measured only once (at the end of the six month period); for these variables there is no way to estimate the initial number of people not knowledgeable nor having the "correct" opinion. These five correlations are not generally stronger than the correlations with the raw average exposures. There appears to be no large change in the magnitude or sign of the correlation either with K equal to 1.5 or with K equal to 2.5.

If the model with the transformations were significantly better than the raw average exposure model, then we would expect to find marked increases in the magnitude

of the correlations for the dependent variables which are most closely related to the theme. The two variables measuring increased knowledge of the function of the United Nations and acquaintanceship with the veto power are those variables which seem most strongly related to the theme; for the first of these variables the correlations does become somewhat stronger but never much stronger than the correlation with the raw average exposures above. For the second variable the correlation becomes weaker. This pattern of little change in the correlations as a result of the transformations is also true for the other themes and time periods. Thus, the model does not seem to add much to the explanation of the changes in the Cincinnati panel over the six months.

In a final attempt to improve upon the transformation model, we have calculated, over the twenty panel variables for which there are measurements both at the beginning and at the end of the six months, an average number of less knowledgeable people for each of the population subgroups. The resulting transformed average exposures (which we have called average transformed average exposures) have also been correlated with the panel variables for K equal to 1.5 and 2.5. These correlations are shown for seven time periods for each of the dependent variables in the table. Again, we find no consistent

increase in the strength of the relationship as the time periods progress and no consistent increase in the strength of the relationship when we compare the average transformed average exposure model and the raw average exposure model.

We conclude that the correlation of the raw average exposure with the panel variables is about as good as with the transformed exposures. Also the correlations at any time period are approximately equal to the correlations at any other time period. Therefore, in Table VIII-22 below for each theme we show the correlations of raw average exposures with each of the panel variables for just one of the time periods. Usually this is the thirteenth time period, but when these data are not available (some data were lost due to computer and/or programming errors) or when they are inaccurate (due to the sizeable cummulation errors in themes 1, 2, 3, and 4), we have felt justified by these findings in presenting correlations for other time periods. We turn now to the presentation and analysis of these data for each of the twelve themes and thirty-one panel variables.

The Correlation Between Exposures to the Themes and the Variables From the NORC Panel

Table VIII-22 shows the correlations over sixteen population subgroups, between the raw average expected

Table VIII-22. Correlations of Raw Average Expected Exposures to the Twelve Scenario Themes with Several Opinion and Information Changes and Recall of Exposure in the Cincinnati Panel over the Six Months Time Period

NORC Panel Measures	1. U.N. Peace-[a] keeping, Promoting Harmony	2. U.S.-[a] USSR Hostility	3. Violence,[a] Wars, Threats to Peace	4. Dissension in the U.N.	5. Anything on the Veto Power	6. Veto Only in Security Council	7. Veto Implies Bi, Power Agreement is Necessary	8. U.N. and Human Rights	9. Cincinnati Plan Sponsors	10. Explicit Explanation of the U.N.	11. Satisfaction with the U.N.	12. Dissatisfaction with the U.N.
General Political Issues and Opinions												
Problem-War	-0.08	-0.07	-0.11	-0.06	-0.08	-0.07	-0.03	-0.02	-0.05	-0.04	-0.07	-0.08
Expect War	0.26	0.32	0.29	0.28	0.16	0.16	0.24	0.16	0.16	0.14	0.15	0.15
Problem-USSR	0.38	0.37	0.38	0.37	0.38	0.37	0.41	0.34	0.31	0.30	0.36	0.35
Count on USSR	-0.02	-0.07	-0.07	-0.01	0.06	0.09	0.04	0.11	0.06	0.09	0.07	0.06
Interest in USSR	-0.12	-0.14	-0.10	-0.12	-0.14	-0.11	-0.15	-0.14	-0.07	-0.10	-0.12	-0.12
Interest in A-Bomb	0.41	0.35	0.36	0.43	0.43	0.46	0.40	0.47	0.47	0.47	0.44	0.44
High Info Change	-0.69*[b]	-0.68*	-0.67*	-0.68*	-0.68*	-0.67*	-0.67*	-0.71*	-0.69*	-0.71*	-0.69*	-0.69*
Medium Info Change	0.70*	0.74*	0.72*	0.68*	0.67*	0.61*	0.65	0.66	0.67*	0.67*	0.65	0.66*
Measures of Exposure to the U.N. Over the Last Six Months												
Media-Newspapers	0.46	0.49	0.49	0.43	0.44	0.40	0.45	0.43	0.41	0.42	0.43	0.42
Media-Radio News	-0.26	-0.24	-0.24	-0.28	-0.26	-0.26	-0.21	-0.30	-0.34	-0.34	-0.27	-0.29
Media-Signs	0.02	0.09	0.04	-0.01	0.01	-0.04	0.01	-0.00	-0.04	0.02	-0.01	-0.02
Media-Leaflets	-0.13	-0.03	-0.09	-0.15	-0.15	-0.20	-0.13	-0.16	-0.20	-0.17	-0.17	-0.16
Media-Meetings	0.57*	0.42	0.50*	0.57*	0.68*	0.74*	0.61*	0.70*	0.72*	0.72*	0.71*	0.69*
Media-Other	0.12	0.18	0.13	0.08	0.13	0.06	0.12	0.11	0.04	0.09	0.10	0.11
Media-Number (3-8)	0.34	0.38	0.35	0.31	0.35	0.31	0.36	0.33	0.28	0.31	0.34	0.33
Knowledge of Various Aspects of the U.N.												
Know U.N. Purpose	-0.31	-0.32	-0.29	-0.31	-0.30	-0.28	-0.28	-0.35	-0.33	-0.35	-0.30	-0.31
U.N. Job-Rights	-0.01	-0.00	0.01	-0.02	-0.00	-0.02	-0.02	-0.01	0.01	0.00	-0.00	0.01
U.N. Job-Health	-0.09	-0.11	-0.13	-0.08	-0.05	-0.03	-0.10	0.03	0.00	0.02	-0.03	-0.05
U.N. Job-Trade	0.18	0.16	0.16	0.15	0.22	0.17	0.14	0.22	0.21	0.24	0.21	0.22
U.N. Job-Bomb	-0.10	-0.09	-0.00	-0.12	-0.05	-0.08	-0.07	-0.07	-0.12	-0.08	-0.07	-0.05
U.N. Veto	0.54*	0.53*	0.52*	0.53*	0.58*	0.56*	0.57*	0.59*	0.56*	0.59*	0.58*	0.59*
Interest in and Attitudes Toward the U.N.												
Interest in U.N.	0.33	0.26	0.30	0.34	0.35	0.38	0.34	0.33	0.36	0.35	0.36	0.37
U.N. Progress (Dissatisfaction with)	-0.40	-0.26	-0.31	-0.39	-0.52*	-0.57*	-0.43	-0.55*	-0.56*	-0.58*	-0.55*	-0.53*
U.N. Prognosis (not succeed)	0.23	0.19	0.25	0.22	0.23	0.26	0.27	0.19	0.21	0.19	0.24	0.21
U.S. U.N. Policy	-0.09	-0.10	-0.10	-0.10	-0.06	-0.04	-0.07	-0.03	-0.06	-0.05	-0.05	-0.09
U.N. Interest Now	-0.05	-0.03	-0.03	-0.09	-0.04	-0.08	-0.01	-0.09	-0.14	-0.11	-0.06	-0.06
More-Publicity	0.20	0.17	0.18	0.20	0.23	0.26	0.24	0.23	0.22	0.22	0.24	0.22
More-Crisis	0.23	0.27	0.24	0.18	0.23	0.17	0.22	0.20	0.15	0.18	0.21	0.21
Know Group	0.03	0.02	0.05	0.03	0.02	0.04	0.08	-0.05	-0.02	-0.04	0.02	0.03
Groups Help	0.19	0.18	0.18	0.18	0.21	0.19	0.22	0.17	0.14	0.15	0.19	0.20
U.N. Slogan	0.26	0.26	0.26	0.24	0.28	0.25	0.31	0.25	0.20	0.21	0.26	0.25

[a]Exposures for themes 1, 2, and 3 are for time periods 9, 7, and 5, respectively, because of the large errors in average exposures for later time periods of these themes. See the text for further discussion of this point. Because of output difficulties with the computer, the final exposures for themes 9, 11, and 12 were lost: for these themes, the average exposures are through time periods 10, 11, and 12, respectively.

[b]Correlations significant at or beyond the 0.05 level are marked with an asterisk.

number of exposures for each of the twelve scenario themes and certain information and opinion changes and recall of exposure variables measured in the Cincinnati panel over the six months by the NORC survey. The most startling and impressive feature of the table is that the largest variation in the correlations is across the panel variables rather than across themes. We might expect that the panel's increasing awarness of problems in dealing with the U.S.S.R. would correlate most highly with the second theme (US-USSR hostility), or at least that the correlation would vary with the theme, but in fact the correlation is nearly constant over the twelve themes.

What can account for this pattern of correlatic ? We know that the pattern of the average number of expected exposures across the sixteen population subgroups is quite similar for each of the themes in the scenario.[15] The main differences between the exposures to the themes is in the overall average level of exposure, which depends upon the number of messages associated with the theme rather than interest in a particular theme.

---

[15]In Table VIII-12 above we noted that males of high school or grade school education and high socio-economic status and females of grade school education and low socio-economic status were the most and least exposed subgroups respectively for each of the twelve themes.

However, those population subgroups which are likely to be exposed relatively highly to one theme are likely to be exposed relatively highly to every other theme in the scenario. Thus, either the messages in the media vehicles are distributed in much the same way from theme to theme or enough duplication exists between the vehicle audiences so that the messages reach the same people regardless of their differences in distribution in the vehicles, or a combination of these occurs.

In general, what are the possible sources of variation in relative exposure of population subgroups across themes? First, if the vehicles or sets of vehicles in the media system reach different relative proportions of the population subgroups, then a concentration of messages in one vehicle or set of vehicles will produce a different distribution of vehicle exposures from that of a concentration of messages in a different vehicle or set of vehicles. One theme might appear mostly in morning weekend radio broadcasts, while another theme might appear primarily in weekday evening newspapers. Table VIII-23 shows the percentage distributions of the three weekday newspapers by sex, age, and education. We find that even though the total audiences differ in size, their distribution across the population types are not greatly different; the fact that a theme occurs predominantly in one or the other of the newspapers would not

change its relative distribution across population subgroups. This is true also for the Sunday newspaper, since its distribution was generated from that of the daily Enquirer. Recall that in our description of the radio vehicle audiences, we used only sex as a defining dimension. Thus, the proportional distribution of these audiences over the other dimensions as produced by the Mosteller parameter estimation routine will be identical. Therefore, the greatest differences which might appear between the vehicle audiences of the two themes would occur if one theme had most of its messages in one of the daily newspapers, while another theme had most of its messages in the radio vehicles. In our scenario it would be very unlikely for one of the themes to have most of its messages distributed throughout the radio vehicles since all of the messages were originally derived from a content analysis of the press. It seems likely, therefore, that the messages of each of the themes were distributed in both the press and the radio and that the resulting distributions of the audiences of the vehicles carrying the messages was nearly equivalent from theme to theme except in the total size.

A second possible cause of different distributions of exposures from theme to theme is possible varying levels of attention to different themes by the population subgroups. These levels of attention are calculated

in the conditional message exposure probabilities of the second stage of the simulation. There are two ways in which these message exposure probabilities can create quite different message audience distributions across the cells even though the vehicle audiences are quite similar.

First, the estimation of the exposure probabilities for men and women from the regression equations might produce for the messages of one theme a quite different distribution of men and women than for a second theme. This could happen if the messages of the one theme were concentrated in the sport pages. Insofar as this is not true, however, the ratio of the average probability of exposure for men to that for women would probably not change significantly from theme to theme. Note also that this regression distinguishes only between men and women and predicts no probabilities based on other of the population-defining dimensions.

Second, the ratios of the conditional exposure probabilities which distribute different probabilities over some of the population dimesnions could be different from theme to theme and would therefore produce significant differences in the message audiences from theme to theme. We might find, for example, that for a given theme the probability of exposure of well-educated people was much lower in comparison with poorly-educated

people than for a second theme. Although the simulation can accept different ratios for each theme, we have not been able to distinguish, within the general area of exposure to international news, thematic differences in the ratios of conditional message exposure probabilities, e.g., for college- and high school-educated people or high school- and grade school-educated people, etc., because these data simply do not exist except for the <u>general</u> category of international news. Thus, this possible source of variation in exposure from theme to theme was not present in our simulation runs.

This is not to say that the distributions of exposures across the population types in the present simulation are incorrect. It may very well be true for this limited variety of news, carried only by newspapers and radio vehicles, that the major differences in exposure from theme to theme are not differences in distribution across population types but rather differences in levels of exposure due to the different numbers of messages and their relative importance as indicated by their format factors. It seems clear, however, that if we were going to specify more precisely the distribution of exposure to the various themes across the population types, we might well invest research dollars investigating additional dimensions of the vehicle audiences of the radio news broadcasts, the different formats of messages and

their effect on message exposure across population types, and, in addition, a content analysis of the radio news broadcasts to determine directly the number and kind of messages carried via these media vehicles.

The finding that those people who are most exposed to one theme are likely to be most exposed to other somewhat similar themes in the mass media seems quite consonant with the common finding of audience studies, that those people who participate highly in one kind of communications behavior participate highly in many other kinds of communication behavior.

The Pattern of Correlations

Although the general pattern of subgroup exposure across themes was constant, perhaps we may find some tendency for the correlations to be higher between panel measures which seem to be more closely related to certain of the themes, implying some slight differentiation in the audiences reached by the themes. Thus, we might expect that the correlations of any of the first four themes with such things as problems or expectations of war or problems with the U.S.S.R. might be slightly higher than the correlations of the other less obviously related themes with these panel variables. In general, this seems not to be the case, although it is true that the correlations of the first four themes with the

increasing expectations of war are somewhat higher in general than that of the other themes. However, this does not hold true for problems with the U.S.S.R., with problems of war, or with the correlations between the knowledge of the veto power in the United Nations and the themes relating to the veto. Thus, we have no evidence in this table that exposure to a particular theme increases conversion or likelihood of change of opinion on a particular NORC panel measure.

From the discussion above we conclude that exposure to any of the themes is equivalent to exposure to international news messages in the mass media system in general, and that this general exposure has different correlations with the different measures in the NORC panel. Which panel variables seem to correlate most highly with this general exposure? The data in Table VIII-22 indicate that exposure to the themes correlates strongly and negatively with a high information change from September to March, very strongly and positively with a medium information change from September to March, strongly and positively with recall of exposure to information about the United Nations in meetings, and also moderately highly with recall of exposure to the United Nations through newspapers over the six months. In addition, exposure seems to correlate highly with increased knowledge of the veto and it correlates

moderately highly with the increasing dissatisfaction with the progress of the United Nations.

Several of the most important changes that occurred in the attitudes of the Cincinnati population over the six months were the large increases in the proportion who named as problems facing the country, another war or maintaining the peace (24 percent to 46 percent), our relations with Russia (16 percent to 29 percent), and also the very large increase in the proportion who expected the United States to enter another war within the next ten years (48 percent to 73 percent). The correlation between the first of these changes and exposure to the mass media is very weak negative correlation for every theme, indicating no overall relationship between increasing concern about war and any of the scenario themes. Thus, if exposure to these themes in the mass media was causing some people to be more aware of the likelihood of war, then there were other influences, e.g., word of mouth messages, which were causing those people in population types not so highly exposed to share the increase in concern about the problem of war.

For the other two changes, there are small positive correlations with exposure, indicating that there was some tendency for those population subgroups who had a higher average exposure to have a higher conversion

index or proportional change in the direction of expecting war or naming the U.S.S.R. as a problem to be contended with. Unfortunately, the thirteenth theme of the simulation containing messages relating to control of the atomic bomb, was lost due to a programming error and therefore we do not have a correlation between this theme and the panel's increasing interest in the atomic bomb. However, we do find a moderatly high correlation between this variable and exposure to the other scenario themes.

One of the most direct relationships likely in this data should be that of simulated exposure and recall of exposure to the United Nations via certain media vehicles. The pattern of high positive correlations for recall of exposure through meetings, newspapers, or through a combination of from three to eight media, and the negative correlations for recall of exposure via the radio or by leaflets may be explained by the plausible assumption that better educated, higher status persons are more likely to be exposed and recall exposure via newspapers, magazines, books, and meetings, and are relatively less likely to be exposed or to recall exposures via radio news broadcasts or through leaflets. In addition, (presumably) these people became increasingly satisfied during the six month period with the progress of the United Nations (the correlations with increasing

dissatisfaction are negative). However, paradoxically, they are increasingly likely to respond that the United Nations probably will not succeed.

How can we explain the strong positive correlation with medium information change and strong negative correlation with high information change? Recall that the index of information change from September to March measures improvement in answering correctly the six questions asking about knowledge of the various aspects of the United Nations. High information change indicates a net increase in the index of from three to five points; medium information change indicates an increase from one to two points. If we assume that the newspaper readers and meeting attenders are also people who are initially relatively well-informed, then they can become only slightly better informed over the six months, resulting in a medium information change. Those people who had a high information change are people who knew very little at the beginning of the six month period (who somehow either learned very much during the period or are guessing and made lucky guesses) and these are people who tend to be little exposed to international news messages in the media system.

One of the interesting effects in these correlations occurs for the question asked of the panel concerning their interest in the United Nations at the end

of the six months compared with their previous interest. Although we find an insignificant negative correlation between exposure and the proportion claiming more interest in the United Nations, the correlations between exposure and the proportion explaining their increased interest either in terms of publicity given to the U.N. or in terms of world crises are both small and positive. Evidently, those people who can rationalize their increased interest are more likely to be exposed.

Finally, we have a small positive correlation between those people who are able to identify the slogan of the U.N. information campaign at the end of six months and those exposed to the mass media.

We may draw several conclusions from our data about exposure to the various scenario themes and its relationship to the NORC panel data. For themes relating to general international affairs news, the data by which we specify the simulation model do not allow us to differentiate very well the patterns of distribution of exposures across population types from theme to theme, except in the overall average level of exposure to each theme, which is primarily a function of the number of messages relating to the theme. In the simulation it appears that the variations in the exposures of the audience across the population types is

governed more by the media habits of the audience than by theme-specific differences in attention.

The correlations in the table may be artificially weak due to random errors in the data. In addition, they may be weak because we have not attempted to simulate several other souuces of exposure, especially the signs, leaflets, meetings, magazines, and films which were used to some degree in the information campaign, nor have we attempted to simulate word-of-mouth or informal communications relating to these themes. The latter could very well be a third stage of the simulation wherein people receiving certain kinds of messages from the mass media, communicate with other people in the simulation population about these messages. Presumably, interpersonal communication often has a stronger effect upon those involved in the communication than messages via the mass media. An interesting question would be what proportion of the variance in conversion or change in the level of knowledge over six months can be explained simply in terms of the messages in the mass media as we have simulated them and what proportion is accounted for by word-of-mouth communications and less formal forms of communications. We note that the strongest correlation found in the present simulation, for recall of exposure to the U.N. in meetings and for medium information change, has a value

of .74. In this case, the exposures to these themes in the mass media as we have modeled them explain at most about fifty percent of the variance in these particular dependent variables. With a simulation model which included various sources of communication, and recorded exposures by source of communication, we might be able to increase the explanatory power and separate the effects of the various sources of communication. This is a possibility for future research.

Finally, we would like to point out that the correlations are across groups and are therefore ecological correlations. We have not correlated individual change with individual exposure because the simulation does not presently report individual exposures (although it does keep internally a record of individual exposures) and also because the survey data is probably not good enough to provide reliable data on individual changes. However, there is one additional reason for not running correlations by individuals. Within a population type, all individuals are equivalent except in their probabilities of exposure, i.e., we have no way initially of associating one individual with an especially high likelihood of conversion and another with a low likelihood of conversion until after the probabilities of being in the audiences of the messages have been computed. Thus, there is no way of making a

one-to-one correspondence, prior to running the simulation, between an individual in the simulation population and a member of the NORC panel.

What can we finally conclude about the implications of the simulated exposures and related changes in the NORC panel for the validity of the simulation? First, although we cannot easily calculate an appropriate correlation between the overall changes in the panel and overall exposure to the simulation themes (because there is no obvious one-to-one matching of panel variables and simulated themes), we do find a high degree of correspondence between those areas of greatest change in the panel and the themes showing the highest average exposure. The largest changes in opinions and attitudes involved the issues of war and peace and relations with the U.S.S.R.: the themes concerned with these issues (the first four themes in the scenario) were by far the most important themes, both in terms of the numbers of actual messages and also in terms of the average number of exposures in the population. The information and opinion areas concerning the functions of the U.N. and levels of support for it, which showed very slight changes in the panel, also had only few messages in the scenario and very low levels of exposure in the simulated population over the six months. Thus, the simulation results for those broad

areas of opinion, attitude, and information seem to support the validity of the model in predicting average levels of exposure for the population.

The second, more difficult, attempt at validation by using correlations across population subgroups, seems not to be so successful. These data do not add significantly to our attempts at validation of the exposures produced by the mass media simulation. Some of the correlations we found here are understandable and others require some strain in interpretation. Our analysis suggests some of the kind of things we can do with this simulation, but it also illustrates the magnitude of the task of attempting a simulation of this sort and the additional kinds of information and data which must be developed in order to make a more specific differentiation of the audiences of such closely related themes.

The Mass Media Simulation: Summary and Conclusions

This study is really a series of studies combined into one. Simulation is apt to force one into this kind of operation because it requires explicit data in many different areas relevant to the problem at hand.

The first stage of the study might be called the elaborate and complete specification of the model. This means that the model is not just a verbal model of the process of exposure in a mass media system, contained in some number of pages of prose, but a mathematical model developing the process from one stage to

another in such detail that it can be programmed for a computer. In the development of this model, of course, there are possibilities for calculation and elaboration due to the fact that the model will be implemented on a computer. However, the computer also imposes constraints in size and time of calculation upon the model.

The next stage in the development is the actual programming of the mathematical model in order to implement it on the computer. In reality, it is not true that the first and second stages are independent of one another; it always turns out that the most complete initial mathematical specification of a very general model such as this omits quite important transitions between various stages or fails to recognize assumptions which must be made and which only come to light in programming the model for the quite literal-minded computer. The present simulation was progra[illegible] and reprogrammed over a period of several years by many different students and finally made operational by Selesnick and Kramer after a year of extensive and intensive work. The result comprises seven different machine loads or links of programs. After programming the model, of course, there must be debugging and test runs and it will happen that contingencies that were not thought of in the previous stages will arise and must be dealt with. Even aft[illegible] his point, however,

when real data are finally applied to the model, there will arise some problems which were not foreseen and necessary adjustments must be made in the model.

Ideally, there is an interaction between the programming and data collection for the model; the model must be specified in terms of parameters which one has some ability to estimate. Somehow this data must be gathered and put in the form required for the computer model. This implies, then, a whole set of studies any one of which might be a rather large task if carried out in rigorous fashion by a single researcher. The specification of the computer population and the audience distribution which is embodied in the first link of the computer model is itself a very time-consuming and tedious task. The available population data is generally not a description of exactly the same population from which the audience data are derived and the discrepancies must somehow be resolved. The audience data is usually incomplete and must be estimated, conjectured, or created in some fashion. The problem of the definition of the actual media as simulation vehicles and the estimation of parameters within this framework complicates the data collection. Finally, relevant to the first stage of the simulation, the division of the population into three distributions for each of the vehicles with cumulation and average audience data for

each and the measurement of duplication between vehicles might well be in itself a long and elaborate study.

Obviously, the content analysis necessary to describe the scenario of messages which appear to the audience could provide at least one thesis if carried out in a most rigorous fashion. This content analysis must not only analyze certain vehicles for occurrences of certain themes, but must also specify such matters as the format of each of the messages carrying the theme. When one is simulating the process of exposure on a computer, it is not sufficient to assume that there is a one-to-one correspondence between appearance of the theme and exposure in a population; therefore, each message is not equivalent to every other message and matters of format, attractiveness, and location of the message become quite important.

Just how important these factors are for readership or listenership is the subject of another kind of study, namely, the effects of format, location, appearance, etc., upon the probabilities of exposure via various channels for various audience types. We have explained in some detail in Chapter VII how such a study might become quite an elaborate affair.

Finally, even after all these studies have been done, the organization of all this data into the framework of the simulation and the processing of survey

type data from the content analysis or the message exposure probability analysis into the kinds of files and data useful for the simulation is itself a tedious, time-consuming undertaking.

We have had problems in evaluating the simulation, because for the best test—comparing actual and simulated exposures to messages (rather than to vehicles)--we must depend upon indirect measures of actual exposure. Even these measures are possibly confounded by the variations in prior exposure and exposures via informal communication channels. We suggest that a better test might involve simulating exposure to information about a new product first being introduced into a community. Nevertheless, the model does succeed in synthesizing consistently the input data and it does produce plausible distributions of exposures. Acknowledging that we are simulating only the portion of exposure due to mass media messages flows, we do feel that the model is valid for this purpose if sufficient attention is paid to the input data.

We would advise the prospective researcher to give some attention to the relative need for precision and comprehensiveness in his input data before he begins to gather it. In simulating with several sources of data, there is always a temptation to be as precise as possible about each data set; however, this can

Table VIII-23. The Relative Distributions of the Average Audiences for the Three Weekday Newspapers, by Sex, Age, and Education

	Male		Female		
	Age		Age		
Education	21-39	40-	21-39	40-	Totals
Enquirer					
College	8.9%[a]	11.2%	4.6%	3.4%	28.1%
High School	12.3	7.6	14.2	11.4	45.5
Grade School	3.0	9.7	1.9	11.8	26.4
Total	24.2%	28.5%	20.7%	26.6%	100.0%
	52.7%		47.3%		
Totals:	21-39 = 44%		40- = 55.1%		
Post					
College	7.1%	4.5%	3.1%	1.2%	15.9%
High School	13.5	8.1	17.5	11.4	50.5
Grade School	6.6	8.8	3.8	14.2	33.4
Total	27.2%	21.4%	24.4%	26.8%	99.8%
	48.6%		51.2%		
Totals:	21-39 = 51.6%		40- = 48.2%		
Times-Star					
College	5.5%	8.5%	4.1%	3.6%	21.7%
High School	10.8	7.8	14.0	11.9	44.5
Grade School	4.1	11.9	3.9	13.9	33.8
Total	20.4%	28.2%	22.0%	29.4%	100.0%
	48.6%		51.4%		
Totals:	21-39 = 42.4%		40- = 57.6%		

[a]The percentages are the percentages of each subgroup in the average audience of the newspaper.

obviously lead to misplaced precision and a misallocation of effort and money. In deciding where to invest his effort, we do not unqualifiedly advocate that the researcher try to match the level of precision of the different data sets, but rather that he try to specify the importance of precision and/or comprehensiveness of each data set <u>for his research goals</u>. Let us illustrate from the present simulation.

In Chapters III through VI, we describe a very large investment of effort in producing the probabilities of vehicle exposure for the Cincinnati mass media system. However, we saw that the content analysis was not reliable (we estimate that twenty-five percent of the relevant messages were missed by the two coders) and, in addition, because of storage problems we were forced to discard twenty percent of the messages in one of the most important themes. We believed, therefore, at first, that much of the effort in specifying the media system was misplaced. Now, however, we will argue that even more effort should have been invested into specifying the vehicle audiences, especially for radio, and perhaps also in relating message exposure probabilities to the content of themes. This is because we feel that the most interesting predictions of

the model relate to the _distribution_[16] of exposure rather than to the overall level of exposure. The former is a function of the vehicle audience distributions, the vehicle distribution of the messages (a relative statistic which could be measured by a small sample of messages), and the ratios of message exposure probabilities. The level of exposure is a function of the total level of vehicle audiences, the number of messages, and the average proportions of the vehicle audiences exposed to a theme (PORTN).

Thus, if we are interested in the changes in the audience distribution from theme to theme, the radio audiences which were differentiated only by size and sex breakdown are prime candidates for further specification. Moreover, since the _formating_ of messages seems to account for only fifty percent of the variance in message exposure probabilities for international news themes, the relationships of these probabilities for the various population types to the _content_ of these themes would seem a worthy area for research. As for the content analysis, the fact that the correlations do not change from time period to time period

---

[16] The knowledge that females of grade school education and low SES are (according to the simulation) ten to fifteen times less exposed than males of high school education and high SES goes very far in explaining the failure of the Cincinnati U.N. information campaign.

and that the same groups are highly or little exposed for themes of from 30 to 1700 messages, indicates that we could have run the simulation through just the first time period instead of thirteen (saving twelve-thirteenths of the seven hours of computer time), or, equivalently, we could have taken only a small sample of messages in the original content analysis, and still have produced the same pattern (not level) of exposures.

For this particular simulation model, we see four changes or additions which would greatly increase its usefulness. First, it seems obvious that frequency distributions as well as averages of expected exposures should be available for the population subgroups. Also the variances of the exposures should be reported. Second, the statistics should be maintained in the computer for such uses as plotting graphs over time and coupling with models of effect and/or word-of-mouth communications.[17] Third, we would prefer that the audience duplication across themes be defined in terms of at least X exposures to theme A and Y exposures to theme B (where the researcher is free to choose X and Y) rather than in terms of at least one exposure, as is presently done. Finally, we are not satisfied with the handling of message exposure probabilities in

---

[17] Except for reporting of the variances, these first two recommendations have already been incorporated into the simulation in the version reprogrammed for the IBM 360-65.

terms of means for the population types. At present, two messages appearing in a vehicle are treated as two separate appearances of the vehicle, each carrying the message. Thus, a person's exposure to the second message is independent of his exposure to the first. Obviously, this is a distortion of reality, just as the assumption of independence in vehicle exposures is a distortion, but in this instance it appears more serious. We might model this better by considering cumulation in message exposures to differentiate message exposure probabilities within the population types.

We have one final admonishment for builders of simulations. Since we always have measurement error and exogenous variables, we should always attempt to describe outcomes with probability distributions, or at least expected values and variances. Also, we should beware of costly Monte Carlo solutions when expected values and variances will save time (money) and provide a more complete description of the likely outcomes.

APPENDIX A

RADIO AUDIENCE MEASUREMENT

One of the sources of radio audience data in this simulation is the Hooper ratings. These ratings are produced by the following process: within a given city, calls are placed to a random sample of the telephone homes located within the non-toll call area of the city. If there is an answer within six rings the respondent is asked the following questions: "Were you listening to your radio just now? ... To what program were you listening, please? ... Over what station is that program coming?"[1]

These calls continue through the day. The rating for a particular program is then the percentage of the homes called which are listening to the program while it is being broadcast.

Nielsen has pointed out several difficulties with this measurement:[2]

1. The sample can only be considered representative of the telephone homes in a city.

---

[1] Hooper, "Introduction," City Hooperatings, p. 2. Additional details of the Hooper coincidental method and trends in the ratings over several years may be found in Mathew N. Chappell and C.E. Hooper, Radio Audience Measurement (New York: American Book-Stratford Press, Inc., 1944).

[2] Arthur C. Nielsen, New Facts About Radio Research (New York: A.C. Nielsen Co., 1946), pp. 26-27.

2. The meaning of the word listening must be decided by each individual as he answers the question. Therefore the quality of listening is variable. (Hooper shows, however, that "... a program's effectiveness in influencing the behavior of listeners is quite independent of any conscious impression."[3])
3. The number of members of the family who are listening is not measured.
4. That member of the family most likely to answer the phone is the one least likely to be listening.
5. Busy signals.
6. Refusals to talk.
7. Deliberate misrepresentations and errors.
8. Some homes have multiple radios and different members of the family may be listening to different programs, thereby confusing the meaning of the measurement.

We will concern ourselves here with the most important problem: how representative of the homes of the Cincinnati Metropolitan District are the telephone homes located in the non-toll call area of the city? According to the

---

[3]Chappell and Hooper, Radio Audience Measurement, p. 103.

telephone company in Cincinnati, approximately eighty-three percent (135,283) of the households in the city of Cincinnati had residential phone service in 1948.[4] An additional 17,412 optional subscribers in Ohio and 31,106 subscribers in northern Kentucky also had toll-free service to Cincinnati phones in 1948. Most of these additional phones were located within the Metropolitan District. The total of 170,380 toll-free subscribers within the Metropolitan District amounted to 62 percent of the households in the area (using the 1950 census estimate of 276,715 households in the Standard Metropolitan Area). Looked at as sixty-two percent of the households, the population represented by the Hooper sample would seem a poor basis for audience estimates for the Metropolitan District. However, we have several good reasons for using the Hooper ratings:

1. Most compelling, they are the only data available.
2. Many other measurements, especially the content of radio news broadcasts, are quite imprecise.
3. The telephone homes did probably represent about eighty-three percent of the households in the non-toll call area. There are no obvious grounds for believing that the telephone homes outside the non-toll call area (but still within the Metropolitan District) were substantially different from those within the non-toll call area. Therefore, we <u>may</u>

[4]These figures and a map of the non-toll call and optional toll-free areas for 1948 were provided by the Cincinnati and Suburban Bell Telephone Company.

have substantially more confidence in the sample than the sixty-two per cent figure would indicate.

The Hooper ratings are coincidental ratings; they measure the average audience of the program, not total audience, since the audience changes from minute to minute. A respondent who listens to only half of a program on each of two occasions would likely be counted in the audience only once, not for both programs. It is the equivalent full-time audience which is measured. This is an underrepresentation of the true audience. The analagous measurement of a newspaper audience would count the person who reads ten per cent of the newspapers as only one-tenth of a person in totaling the audience! Nielsen, on the basis of electronically received (Audimeter) ratings, estimates that the average audience measurement cuts the ratings of stations and programs by twenty to forty per cent.[5] Of course, the longer the programs, the more likelihood of changes in the audience, and therefore the greater the difference between the total audience and the coincidental audience. The chart below (Figure A-1) demonstrates quite clearly the effect of program length.[6]

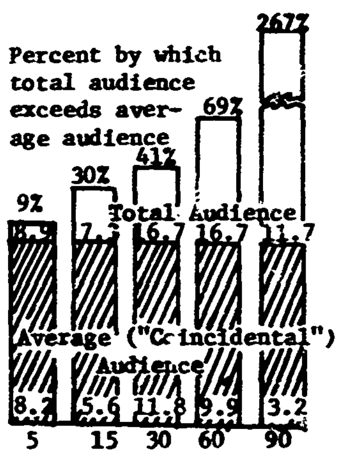


Figure A-1. Total Audience vs. Average ("Coincidental") Audience.

[5]Nielsen, New Facts, p. 46.

[6]The chart is from Nielsen, New Facts, p. 52. Although the dates are not provided, the data probably came from 1945 Audimeter measurements.

Other Nielsen data, represented in the chart below (Fig. A-2), <u>seem</u> to show the effect of type of program upon the difference between the total and average audience.[7] The data, given in terms of the percentage of the total program time listened to by the average audience, demonstrate that news broadcasts rank second only to the daytime serials in holding the audience. However, we must be cautious in attributing the differences in holding power to the program types, since it seems likely that they fall into an order of length, i.e., the daytime serials and news broadcasts which seem to have the greatest holding power tend to be fifteen minutes in length, while at the other extreme the concert music broadcasts which have the poorest holding power are probably also the lengthiest programs.

Since the amount of international news per broadcast is usually about three continuous minutes,[8] we expect from the Nielsen figures that the total audience exposed to it is very nearly equal to the average audience

---

[7]This Nielsen data is cited in C.H. Sandage, <u>Radio Advertising for Retailers</u> (Cambridge: Harvard University Press, 1945), pp. 142-143. No date is given.

[8]See the results of Xavier University Study cited above, p. 121.

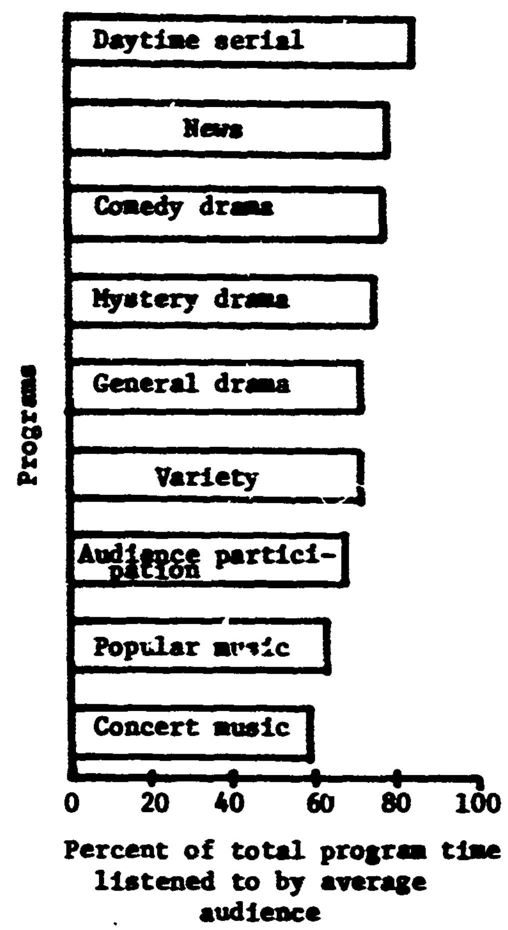


Figure A-2. Holding Power Variations in Different Types of Programs. (Source: Nielsen Radio Index, A.C. Nielsen Company, Chicago)

APPENDIX B

A PROBABILISTIC MODEL OF EXPOSURE

Assign a number to each member of the population from 1 to K, e.g., $j = 1, 2, 3, \ldots, K$.

Define the event A_i^j:

$A_i^j \equiv$ the jth member of the population is exposed at least once during the ith day.

Define the random variable X_i^j and associate with it the probability P_j in the following manner.

$$X_i^j \equiv \begin{cases} 1 \text{ if the event } A_i^j \text{ is observed, and this happens with probability } P_j. \\ 0 \text{ otherwise, with probability } 1-P_j. \end{cases}$$

P_j is a number between zero and one such that if one observes over n days, then in the limit as n becomes very large, P_j represents the frequency of exposure of the jth person:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^j \doteq P_j \quad \text{or} \quad \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^j = P_j$$

In this first simple model we assume that the probabilities P_j are constant over time and independent from person to person and from day to day. Thus the process of the jth person being exposed on the ith day is a Bernoulli process.

The probability P_j is the probability of the event A_i^j. But we also assume that it is characteristic of the jth person in the same way that a coin has a characteristic probability of heads. It attaches itself to the jth person in that it represents the likelihood per day of his being exposed via the particular vehicle in question. When we attempt a computer simulation of the process of exposure of a population via the vehicle, we store as data a list of these probabilities, one probability for each member of the population.

Below we shall define a random variable π_r^n whose value is the number of people in the population who are exposed on exactly r of n days. We now make the assumption that the unknowns in the model (the values of K and the P_j's) are so related that the expected values of the random variable π_r^n for $r = 0,1,2,\ldots,n$ are "very close" to the observed values of the frequency curve.[1] This assumption indicates how the model is related to the observed data and implies some relations among the unknowns; however it falls far short of completely specifying the unknowns (even how many there are) and this task requires a much stronger assumption, namely the form of the distribution of the P_j's (to be discussed below).

---

[1] I think we should ideally assume that the unknowns are so related as to make the most likely values of π_r^n, $r = 0,1,2,\ldots,n$ "very close" to the observed values of the frequency curve. However, as we shall see below it is practically

Now we show the implications of the model. If we define a random variable r_j representing the number of times out of n that the jth person is exposed:

$$r_j = \sum_{i=1}^{n} X_i^j$$

then r_j is just the sum of n Bernoulli trials and its probability is the binomial probability. (We drop the j on the r here because it is carried on the P_j's.) The probability that the jth person will be exposed on exactly r of n days, given P_j, is

$$P_j'(r/n, P_j) = \binom{n}{r} P_j^r (1-P_j)^{n-r}.$$

Now let us define the new random variable:

$$Y_j = \begin{cases} 1 & \text{if the jth person is exposed on exactly r of n days, with probability } P_j' \\ 0 & \text{otherwise, with probability } 1-P_j'. \end{cases}$$

Then the number of people in the population who are exposed on exactly r of n days is given by the random variable:

$$\Pi_r^n = \sum_{j=1}^{K} Y_j .$$

---

impossible to calculate the joint likelihood function for the Π_r^n and extremely easy to calculate the expectations. Probably both assumptions imply very nearly the same relationships among the unknowns.

It is very difficult to write the probability distribution for π_r^n. π_r^n can take on values from 0 to K. For each of these values the associated probability consists of a sum of $\binom{K}{value}$ terms, where each term is the product of K probabilities, the P_j''s and/or $1-P_j'$'s. However, we can easily write the expected value of π_r^n using the rule that the expected value of a sum is equal to the sum of expected values of the terms. Thus:

$$E(\pi_r^n) = E(Y_1) + E(Y_2) + \ldots + E(Y_K).$$

Now in general (from above):

$$E(Y_j) = P_j' \cdot 1 + (1 - P_j') \cdot 0 .$$

Therefore:

$$E(\pi_r^n) = \sum_{j=1}^{K} P_j' = \sum_{j=1}^{K} \binom{n}{r} P_j^r (1 - P_j)^{n-r} \qquad \text{(B-1).}$$

This is the formula for the expected number of people who are exposed exactly r days in n.

Now if we had (for example only) people grouped by probability into h groups, of probability P_k and number of persons n_k, then the sum would become:

$$E(\pi_r^n) = \sum_{k=1}^{h} n_k \binom{n}{r} P_k^r (1 - P_k)^{n-r} \quad \text{where} \quad \sum_{k=1}^{h} n_k = K \qquad \text{(B-2).}$$

If we wish to consider the continuous case, we can do so as follows:

In place of n_k we let $N(p)$ be a population density function such that $N(p)dp$ = the number of people with probability between p and p+dp (this corresponds to P_k) and such that

$$\int_0^1 N(p)dp = K$$

Then the sum becomes an integral and the formula for the expected number of people who are exposed on exactly r of n days is given by:

$$E(\pi_r^n) = \int_0^1 N(p)\binom{n}{r} P^r (1-P)^{n-r} dp \qquad \text{(B-3)}.$$

We can easily return to our first form by letting the population density function be a sum of Dirac delta functions,

$$N(p) = \sum_{j=1}^{K} \delta(P-P_j).$$

then,

$$E(\pi_r^n) = \int_0^1 \sum_{j=1}^{K} \delta(P-P_j)\binom{n}{r} P^r (1-P)^{n-r} dp$$

$$= \sum_{j=1}^{K} \binom{n}{r} P_j^r (1-P_j)^{n-r}$$

We can also derive the formula for the variance of π_r^n. Since the Y_j are independent, the variance of

the sum is equal to the sum of the variances. Thus

$$Var(\pi_r^n) = Var(Y_1) + Var(Y_2) + \ldots + Var(Y_K)$$

The variance of Y_j is just

$$Var(Y_j) = P_j'(1-P_j') \ ;$$

therefore

$$Var(\pi_r^n) = \sum_{j=1}^{K} P_j'(1-P_j') = \sum_{j=1}^{K} \binom{n}{r} P_j^r (1-P_j)^{n-r} \left[1-\binom{n}{r} P_j^r (1-P_j)^{n-r}\right].$$

for the continuous case, the variance of π_r^n becomes

$$Var(\pi_r^n) = \int_0^1 N(p) \binom{n}{r} P^r (1-P)^{n-r} \left[1-\binom{n}{r} P^r (1-P)^{n-r}\right] dp \quad \text{(B-4)}.$$

A second quite valuable statistic is the cumulative number of people exposed at lease once in t time periods. We define a new random variable:

$$Z_j = \begin{cases} \text{1 if the jth person is exposed at least once} \\ \text{in t days, with probability } \left[1-(1-P_j)^t\right] \\ \text{0 otherwise, with probability } \left[(1-P_j)^t\right] \end{cases}$$

Then the total number of people in the population who are exposed at least once in t days (the t-period cumulative audience) is given by the random variable

$$C_t = \sum_{j=1}^{K} Z_j$$

Again, it is very difficult to write the probability

distribution for C_t: however we can easily calculate the expected value and variance of the statistic just as above for π^n_r:

$$E(C_t) = \sum_{j=1}^{K} E(Z_j) \qquad \text{and} \qquad Var(C_t) = \sum_{j=1}^{K} Var(Z_j).$$

Therefore

$$E(C_t) = \sum_{j=1}^{K} \left[1 - (1 - P_j)^t\right]$$

and

$$Var(C_t) = \sum_{j=1}^{K} \left[1 - (1 - P_j)^t\right]\left[(1 - P_j)^t\right].$$

For the case of the continuous population density function:

$$E(C_t) = \int_0^1 N(p)\left[1 - (1 - P)^t\right] dp \qquad \text{(B-5).}$$

$$Var(C_t) = \int_0^1 N(p)\left[1 - (1 - P)^t\right]\left[(1 - P)^t\right] dp \qquad \text{(B-6).}$$

For the continuous case Hyett[2] has suggested using a beta density function to model the probability distribution

---

[2] G. P. Hyett, paper read to the Statistics Seminar, London School of Economics, February, 1958.

and is reported to have tested it successfully on data from American magazines. Below we outline the mathematics of the single beta function (which we can then extend to the three beta function case).

The Single Beta Function Model

The beta function of p is defined over the interval $0 \leq p \leq 1$ as[3]

$$f_\beta(P/m',n') \equiv \frac{P^{m'-1}(1-P)^{n'-1}}{\beta(m',n')}$$

where m' and n' are two parameters, each greater than zero, and β(m',n') is a constant chosen such that

$$\int_0^1 f_\beta(P/m',n')\,dp = 1.$$

This function is often quite useful in representing probabilities since it can take on a rich variety of forms over the interval depending upon the choice of the two parameters. These forms can be as diverse as a U-shape, a horizontal line, or a bell shape, and can be symmetric or skewed toward either end of the range

The mean of the function is given by

$$\bar{P} = \frac{m'}{m'+n'}$$

---

[3] This notation is read "the beta function of p given m' and n'."

and the variance by

$$\check{p} = \bar{p}(1-\bar{p})\frac{1}{m'+n'+1} .$$

Also the maximum of the function occurs at

$$p = \frac{m'-1}{m'+n'-2}$$

except for the cases $n' \leq 1$ and/or $m' \leq 1$ in which cases the maximum occurs at the points p=0 and/or p=1. The next page shows plots of the function for several values of m' and n'.[4]

It can be shown[5] that

$$\beta(m', n') = \frac{\Gamma(m')\,\Gamma(n')}{\Gamma(m'+n')} \tag{B-7}$$

where

$$\Gamma(x+1) = x\,\Gamma(x) \tag{B-8}.$$

Multiplying the beta function probability distribution by the population size, K, gives the population density function

$$N(p) = K f_\beta(p/m', n').$$

Then N(p)dp is the number of people with probability between p and p+pd and the total population is just

$$\int_0^1 N(p)dp = K \equiv \text{the total population.}$$

---

[4] For a more detailed discussion of the beta function see Howard Raiffa and Robert Schlaiffer, Introduction to Statistical Decision Theory (New York: McGraw-Hill, 1965).

[5] Ibid.

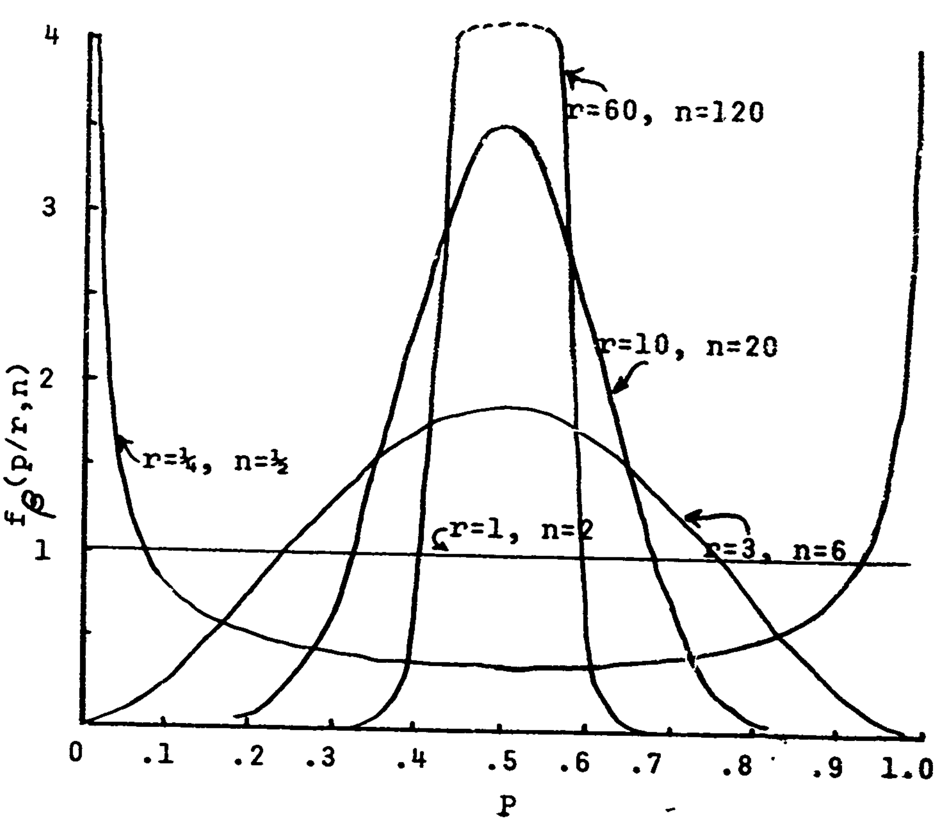


Figure B-1. Beta Densities, $\overline{P}$ = 0.5

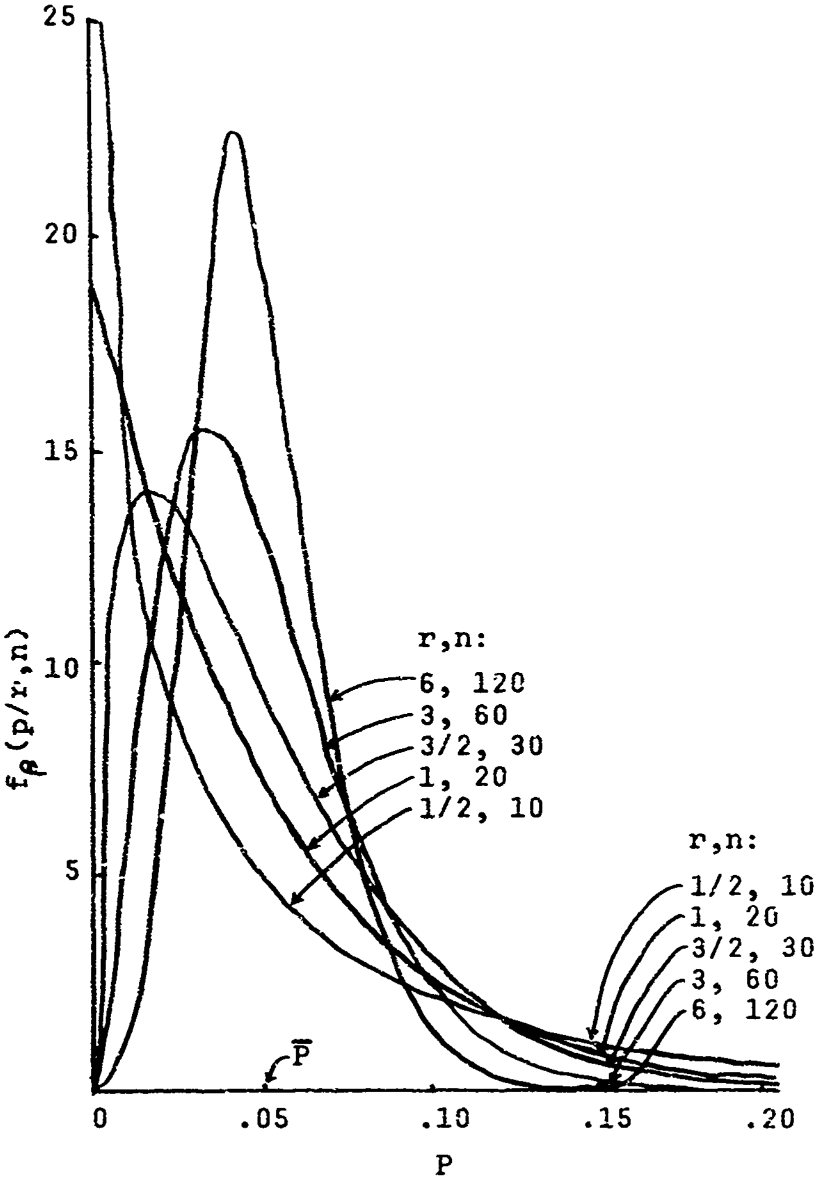


Figure B-2. Beta Densities, $\overline{P}$ = .05

Now with this function the expected number of people exposed at least once through t time periods follows from equation (B-5) above,

$$E(C_t) = \int_0^1 N(p)\left[1-(1-p)^t\right]dp$$

$$= \int_0^1 K f_\beta(p/m', n')\left[1-(1-p)^t\right]dp$$

$$= \frac{K}{\beta(m', n')}\int_0^1 p^{m'-1}(1-p)^{n'-1}\left[1-(1-p)^t\right]dp$$

$$= \frac{K}{\beta(m', n')}\left[\beta(m', n') - \beta(m', n'+t)\right]$$

$$= K\left[1 - \frac{\beta(m', n'+t)}{\beta(m', n')}\right].$$

Using the relationships between the complete beta function, the gamma function, and factorials (equations (B-7) and (B-8)) we simplify the result

$$E(C_t) = K\left[1 - \frac{\Gamma(m')\,\Gamma(m'+n')\,\Gamma(n'+t)}{\Gamma(m')\,\Gamma(n')\,\Gamma(m'+n'+t)}\right]$$

$$= K\left[1 - \frac{(n'+t-1)(n'+t-2)\cdots n'}{(m'+n'+t-1)(m'+n'+t-2)\cdots(m'+n')}\right] \quad \text{(B-9).}$$

This formula provides the expected cumulation through any number of periods, if the values of n' and m' are known.

Typically our data will consist of measurements or estimates of the average audience (C_1) and the two-period cumulation (C_2). Given these values we can solve for the two parameters of the beta function. From (B-9) with t = 1 and t = 2

$$E(C_1) = K\left[1 - \frac{n'}{m'+n'}\right] \qquad (t=1)$$

$$E(C_2) = K\left[1 - \frac{(n'+1)n'}{(m'+n'+1)(m'+n')}\right] \qquad (t=2).$$

Since our purpose is to equate these expected values with the measured or estimated values

$$E(C_1) = C_1$$
$$E(C_2) = C_2$$

and therefore

$$C_1 = K\left[1 - \frac{n'}{n'+m'}\right]$$

$$C_2 = K\left[1 - \frac{(n'+1)n'}{(m'+n'+1)(m'+n')}\right].$$

Solving these two equations for the values of n' and m' in terms of C_1 and C_2 gives

$$n' = \frac{(K-C_1)(C_1-C_2)}{KC_2-2KC_1+C_1^2} \qquad \text{(B-10)}$$

and

$$m' = \frac{C_1}{K-C_1}n' = \frac{C_1(C_1-C_2)}{KC_2-2KC_1+C_1^2} \qquad \text{(B-11)}.$$

Equations (B-10) and (B-11) enable the calculation of the values of the parameters from C_1 and C_2. However they also imply certain limitations on the values of C_1 and C_2 which can be modeled by the distribution. Since n' must be greater than zero (in order that the integrals converge) it follows from equation (B-10) that

given $K > C_1$, i.e., the total propulation greater than the average audience and

$C_2 > C_1$, i.e., the two-period accumulation greater than the average audience,

then the denominator of equation (B-10) must be negative if n' is to be greater than zero,

$$KC_2 - 2KC_1 + C_1^2 < 0$$

or

$$\frac{C_2}{K} < 2\frac{C_1}{K} - \frac{C_1^2}{K^2} \qquad \text{(B-12)}.$$

In words, this equation implies that the mean proportion of the population exposed through two periods must be less than twice the average audience proportion. The equality of the left and right hand sides of equation

(B-12) would imply a random process with each individual having the same probability of exposure, i.e., the entire population lumped at the <u>same</u> probability. Equation (B-12) implies that this condition is the limiting form which can be modeled by the beta density function.

The variance of these expected cumulations is found from equation (B-6)

$$Var(C_t) = \int_0^1 N(p)\left[1-(1-p)^t\right](1-p)^t \, dp$$

$$= K\left[\frac{\beta(m_i', n_i'+t) - \beta(m_i', n_i'+2t)}{\beta(m_i', n_i')}\right] \qquad \text{(B-13).}$$

We also calculate the mean and variance of the π_r^n. Using the same population density function the expected number of people exposed on exactly r of n days is (from equation (B-3)):

$$E(\pi_r^n) = \int_0^1 N(p)\binom{n}{r} p^r (1-p)^{n-r} \, dp$$

$$= \int_0^1 K f_\beta(p/m', n')\binom{n}{r} p^r (1-p)^{n-r} \, dp$$

$$= \int_0^1 \binom{n}{r}\frac{K}{\beta(m', n')} p^{m'+r-1}(1-p)^{n'+n-r-1} \, dp$$

$$= \binom{n}{r} K \frac{\beta(m'+r, n'+n-r)}{\beta(m', n')} .$$

This can be "simplified" by use of the relations in equations (B-7) and (B-8):

$$E(\pi_r^n) = \binom{n}{r} K \frac{\Gamma(m'+r)\Gamma(n'+n-r)\Gamma(m'+n')}{\Gamma(m'+n'+n)\Gamma(m')\Gamma(n')}$$

$$= \binom{n}{r} K \frac{(m'+r-1)(m'+r-2)\cdots m'(n'+n-r-1)(n'+n-r-2)\cdots n'}{(m'+n'+n-1)(m'+n'+n-2)\cdots(m'+n')}$$

(B-14).

The variance of π_r^n is calculated from equation (B-4),

$$Var(\pi_r^n) = \int_0^1 N(p)\binom{n}{r} p^r (1-p)^{n-r}\left[1-\binom{n}{r}p^r(1-p)^{n-r}\right] dp$$

$$= K\binom{n}{r}\left[\frac{\beta(m'+r, n'+n-r) - \binom{n}{r}\beta(m'+2r, n'+2n-2r)}{\beta(m', n')}\right]$$

(B-15).

Summary of the One Beta Function Model

The population is distributed as a function of the probability according to

$$N(p) = K f_\beta(P/m', n')$$

where K equals the total population, and m' and n' are determined by fitting known values of C_1 and C_2. These values must satisfy the following conditions

$$K > C_2 > C_1 \qquad \text{(B-16)}$$

and

$$\frac{C_2}{K} < \left(\frac{C_1}{K}\right)\cdot\left(2 - \frac{C_1}{K}\right) \qquad \text{(B-17).}$$

The Three Beta Function Model

The population is considered to consist of three groups, the frequent, infrequent, and casual audience of a vehicle. For the case of a magazine, the frequent readers might be the subscribers and their families, the casual readers might be the newsstand buyers and passalong readers, and the infrequent readers the rest of the population. It will be typical of this approach that the values of some parameters will be residual, as well as the estimate of the number of people in one of the three groups. This is because the best empirical values will usually apply to the total population. The following equations summarize the obvious relationships between the parameters and population values for each of the three groups and the totals for the entire population. (The Roman superscripts on the C's indicate the group to which they refer. Those without Roman

superscripts refer to the total population.) First the contributions to the cumulation from each of the distribution must sum to the total for the whole population. Therefore

$$C_1^{i} + C_1^{ii} + C_1^{iii} = C_1 \qquad \text{(B-18)}$$

$$C_2^{i} + C_2^{ii} + C_2^{iii} = C_2 \qquad \text{(B-19)}$$

Also the population values of the three groups must sum to the total, i.e

$$K^{i} + K^{ii} + K^{iii} = K \qquad \text{(B-20)}.$$

In general the values of C_1, C_2, and K will be the best estimates. Therefore for each equation it suffices to specify only two of the values for the subgroups. In the example below the number of infrequent readers is calculated as a residual. Using equation (B-20)

$$K^{i} = K - K^{ii} - K^{iii}$$

or in words,

$$\begin{pmatrix}\text{number of}\\ \text{infrequent}\\ \text{readers}\end{pmatrix} = \begin{pmatrix}\text{total}\\ \text{population}\end{pmatrix} - \begin{pmatrix}\text{newsstand}\\ \text{buyers}\end{pmatrix} - \begin{pmatrix}\text{subscribers}\\ \text{and their}\\ \text{families}\end{pmatrix}$$

The problem of Chapter III is just to make these estimates of the best values for the K's, C_1's and C_2's for the various media. Of course the restrictions of (B-6) and (B-7) apply to each set of $[K, C_1, C_2]^{i}$.

The audience data for the Cincinnati vehicles of 1947-1948 is not sufficient to adequately test the beta function model. For none of the vehicles do we have any data beyond a good estimate of the average audience; for the two-period cumulations we have made rough estimates on the basis of these average audiences and educated guesses about distributions and their mean exposure probabilities (see Chapter V for details). Therefore we must turn to data from other vehicles and media to assess the usefulness of the beta function model.

The Fit of the Beta Function Model to Empirical Data

As we have pointed out in the Introduction, the beta function model appears to have been first suggested by Hyett in a seminar at the London School of Economics.[6] Hyett proposed using one beta function for the total population and is reported (by Metheringham) to have tested the model on empirical data with favorable results. Metheringham extended Hyett's calculations to duplication between vehicles in an article in the *Journal of Advertising Research*.[7]

---

[6] G.P. Hyett. Paper read to the Statistics Seminar, London School of Economics, February, 1958.

[7] Richard A. Metheringham "Measuring the Net Cumulative Coverage of a Print Campaign," *Journal of Advertising Research*, Vol. 4, No. 4 (December, 1964), pp. 23-28.

These previous attempts have taken a single beta function and used the average and the two-period cumulative audience to fit the two parameters of the beta function. However, this method in general does not provide the best estimate of the two parameters of the beta function, especially if additional cumulation data are available. In this case, we could choose the values of the parameters which maximize the probability of getting the observed cumulation curve (maximum likelihood estimation), or combine the likelihood function with a prior estimate of the parameters to get a Bayesian estimate. A third possibility would be to choose the parameters so as to minimize the summed, squared deviations between the observed cumulation curve, and that implied by the beta function. Note however that the beta function distribution of probabilities implies an infinite set of frequency and cumulation curves. Thus one least squares procedure would attempt to equate the average or "expected" cumulation curve with the observed curve. This is equivalent to the procedure outlined above (p. 474) when only the first two cumulation values are available (i.e., fitting C_1 and C_2). Since these values are generally the only values available (at best) to the researcher attempting to simulate a mass media system, we have chosen this estimation procedure and shall use it in the following exploration of the fit of model to empirical data.

There are some complications to this method of estimation. First, let us note that it seems unreasonable that every point in the cumulation curve should carry equal weight. For example, the cumulative two-period audience (C_2) includes the audience from C_1 (the average audience). In no sense can we imagine that the second point (C_2) on the curve is independent on the first point. This holds true for all the rest of the points on the curve. For this reason, it would seem reasonable to assign decreasing weight to the values of the observations as the time period increases. This would be the reasonable way to allot weight to the empirical data; however, it does not necessarily produce the closest fit between the derived and the empirical points, and that closest fit is obviously the least squares solution for the two parameters.

A third reason for using only C_1 and C_2 in fitting the curves is that the derived curves are very complicated functions of the parameters, and the mathematics involved in a least squares estimation of the parameters from all the points on the curves is extremely difficult and probably impossible, except in some iterative technique on a computer. For all these reasons, then, we have followed the conventional procedure of fitting the empirical curves from values of C_1 and C_2 even though this does not necessarily provide the best fit to the data over the entire length of the curve. In general, it will not be the best possible fit to the data over the entire length of the curve.

In the single beta function model there are two parameters to be estimated from the data. However, in the three beta function model, we have a total of eight parameters which must be estimated from the data. In this case, for each distribution we must estimate C_1 and C_2 and for two of the distributions we must estimate the proportions of the population belonging to the distributions. The third proportion, of course, is the proportion needed to make the entire population sum to 100 per cent. Given the difficulty in getting empirical values for the one beta function model, one might imagine that it is doubly difficult to get empirical values for the eight parameters of the three beta function model. (For an example of the kinds of reasoning used in generating these parameters and the relationships between these parameters, see Chapter V.) Because of the imprecision in the estimates of these parameters, we have made some attempt to examine the problem of what difference the parameters make in values of cumulation and frequency of exposure. To do this, we have taken as given two values, the average audience and the two-period cumulative audience, for the entire population. This completely specifies the parameters for the one beta function model, but it leaves six of the eight parameters free in the three beta function model. The question we ask is: given that the cumulation and average audience are fixed for the first two time periods

for the entire population, what difference does juggling the other parameters make in the values of cumulation and frequency of exposure? To answer this we have taken three cases for exploration. This is actually only an exploration because it would be very difficult to examine all of the possibilities, the infinite number of combinations of the six remaining free parameters, even though they are somewhat constrained by each other. In two of the three explorations we have a great deal of empirical data available with which to compare the one beta function and three beta function models. In the last case we have taken hypothetical values for the average audience and the two-period combination which might be appropriate to a newspaper and explored the consequences of various combinations of the remaining six parameters in the three beta function model.

Instead of an exhaustive exploration, what we have done is choose values for the parameters which seem plausible in the light of data or intuition about the distribution of infrequent and quite frequent readers in the population, looking at the effects that this variation in parameters has upon the frequencies and accumulations generated. In analyzing this data we may keep two points in mind. The first of these is that for each distribution, the inequality constraint relating the maximum value of the two-period cumulation to the average audience limits the possible values of that two-period cumulation, given the average audience. The second point to remember is that the single

beta function model is equivalent to a three beta function model when the average and two-period cumulative audience as proportions of the population of each distribution are identical with these proportions as found in the population as a whole. Keeping these facts in mind, let us look at the first exploration of the parameters.

For the first case there is a great deal of data available. This is a study of the cumulative and repeat audiences of Better Homes and Gardens over twelve issues, beginning in October, 1954 and continuing through September, 1955.[8] For Better Homes and Gardens we chose the following values for the parameters; in each trial we set the proportion of the population in the low distribution at 20 percent with an average probability of exposure of .01 and a proportional two-period cumulation of .019. The values for the high distribution range from two-and-a-half to ten percent of the population in the high distribution with an average exposure probability ranging from .80 to .95 and the proportion of the population in the two-period cumulative audience ranging from .93 to .99. The values for the middle distribution are found by subtraction from the overall average exposure probability, which is .125, and the overall average proportion in the two-period cumulative audience, which is .182. With these

---

[8] This data is taken from Alfred Politz Research, Inc., A Twelve Months' Study of Better Homes and Gardens Readers, Des Moines, Iowa: Meredith Publishing Company, 1956.

ranges of values, we look at several of the expected cumulation and frequency statistics derived from the resulting distributions (Fig. B-3). It is obvious from the graph that one of the cumulative curves deviates rather markedly from the empirical cumulation curve and also the frequency curve in this case is quite different from the empirical curve. Both the single beta function curve and two of the three beta function curves fit the empirical data rather well. The empirical twelve-period cumulative audience is .357. The one distribution model gives a predicted twelve-period cumulative proportion of .342, somewhat lower than the empirical value. For the three beta function case, however, the values generated encompass the empirical value; they range from a minimum proportion of .345 to an upper limit of .483. Thus, the three beta function model can cover the empirical values; after twelve issues it has a variation of thirteen and one-half per cent of the population or, at most, it over-estimates the twelve-period cumulative audience by about 30 percent.

Next we look at several of the frequency values which seem to be important. The usual frequency curve shows a large number of people exposed none out of N times, somewhat fewer exposed once, and in general, the curve declines towards exposure about half of the time, then increases somewhat for exposure to nearly all of the issues, due to the subscribers in the population.

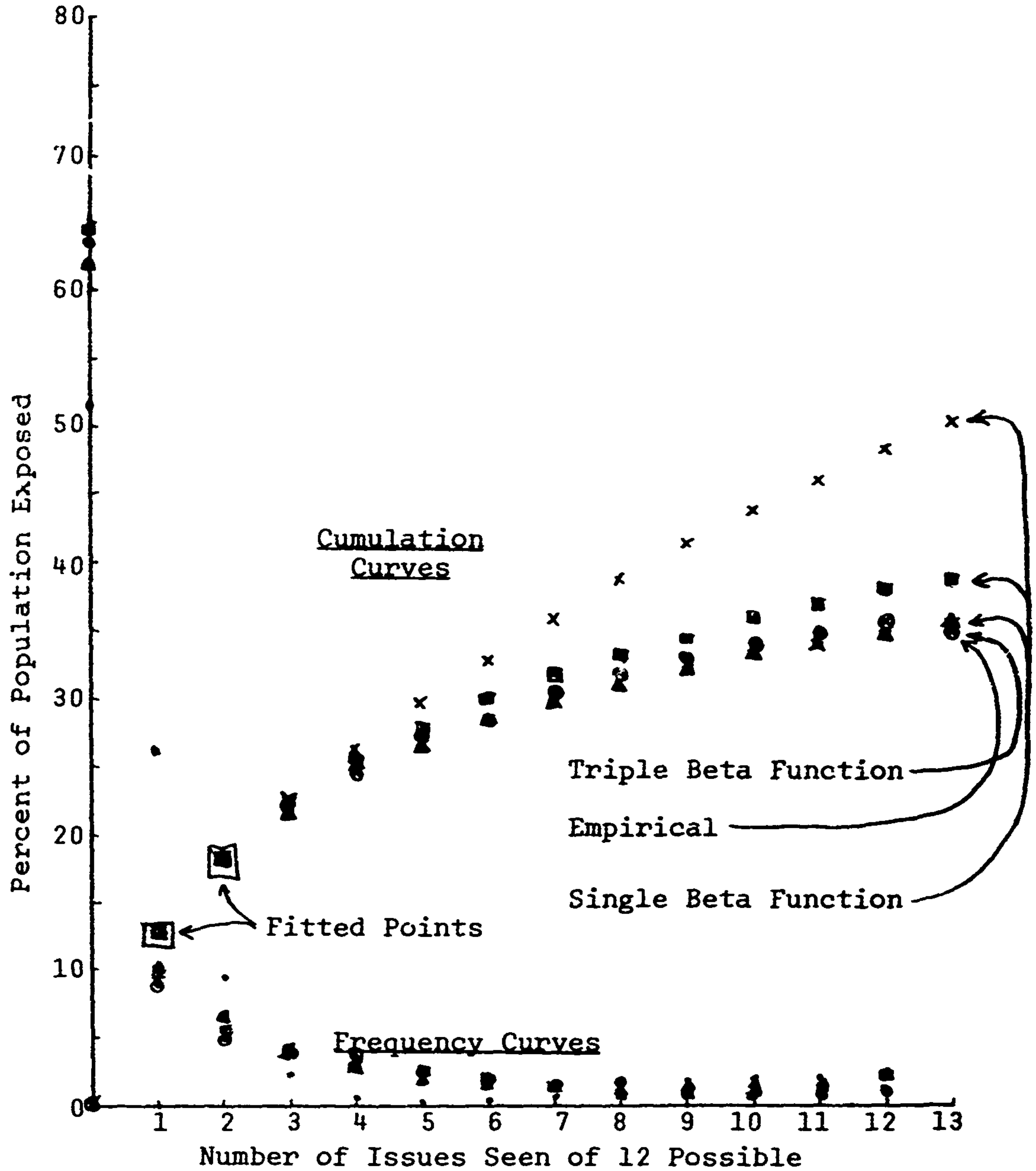


Figure B-3. Cumulation and Frequency Curves For Various Values of the Parameters For <u>Better Homes And Gardens</u>, 1958

If we look at the frequency curve corresponding to the most deviant cumulation curve we see that the frequency curve shows many more people with very infrequent exposure, much fewer than average numbers of people with moderate exposure and then a slight growth in the number of people with high exposure. Thus the effect which gives such a high rate of cumulation is to increase the number of people at each end of the frequency curve and decrease the proportion of cases in the middle ranges of the curve.

The second magazine for which we have empirical data and have generated a number of cumulation and frequency curves is a Swedish magazine, Hemmets Veckotidning, which was studied by Schyberger in Malmo during 1962.[9] For H. V., the average audience was 33.4 percent of the population with an average two-period cumulation of 40.5 percent of the population. Thirteen different curves were fitted with the following values of the parameters: the proportion of the population in the low distribution varied from ten to forty percent while the population proportion for the high distribution varied from ten to twenty-five percent. The average probability for the low

---

[9] The data on this magazine and on several other Swedish magazines is reported by Bo W:son Schyberger in Methods of Readership Research, pp.93-117. We note here that the general method of averaging over all the issues studied in order to get the average audience and the average cumulation figures was not used by Schyberger in his reporting. Therefore, his values of two-period cumulation are not the values which one

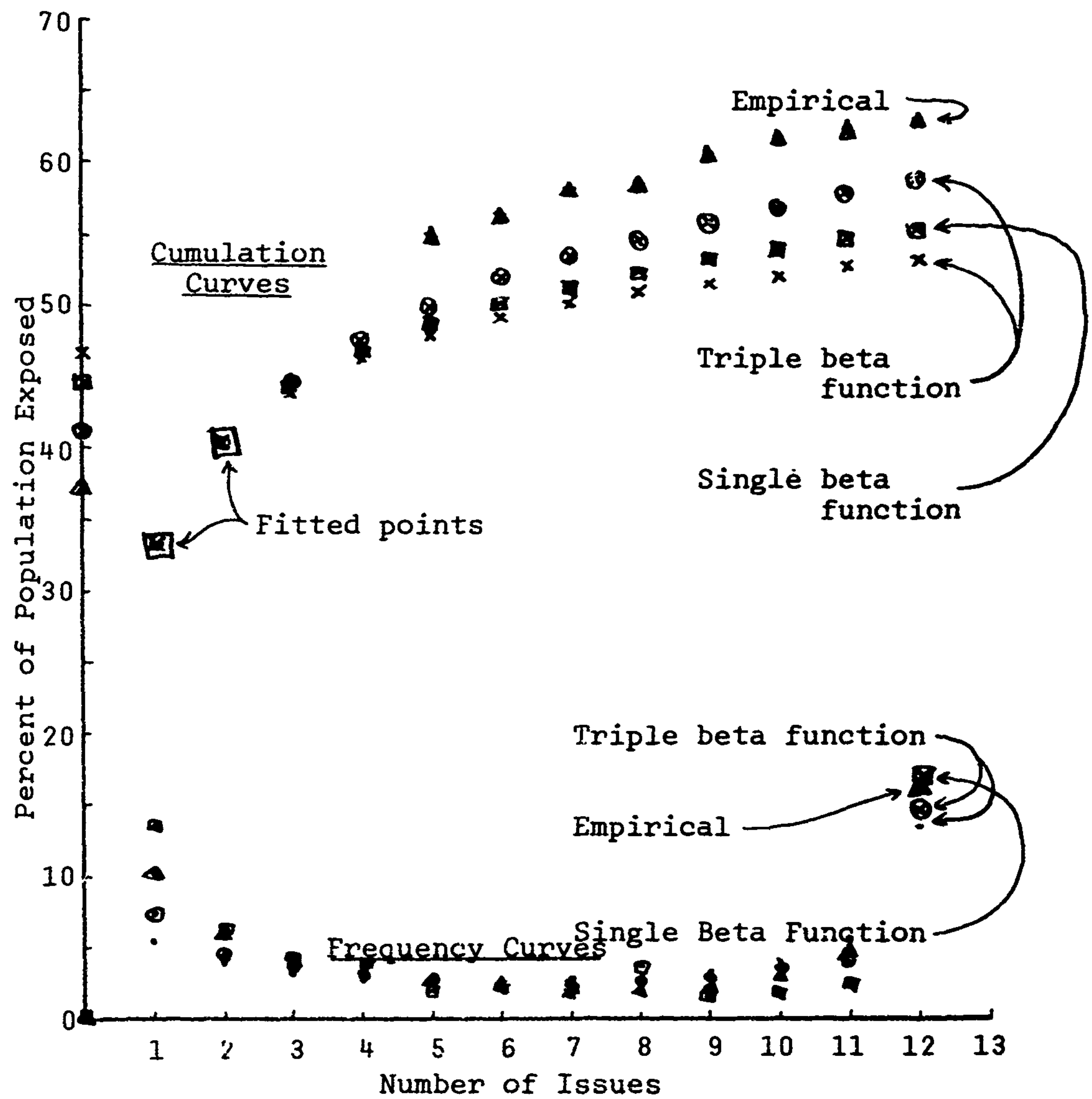


Figure B-4. Cumulation and Frequency Curves for Various Values of the Parameters For H.V. (Swedish Magazine)

distribution varied from .010 to .100 with a cumulation varying from .015 to .170. For the high distribution the average probability varied from .800 to .900 with cumulation varying from .900 to .980. The particular cumulation curve which comes closest to the empirical curve was generated with forty percent of the population in the low distribution with a mean probability of .100 and a two-period cumulation of .170. The proportion of the population at the high distribution was ten percent with a mean probability of .900 and two-period cumulation of .970. This left fifty percent of the population in the middle distribution with a mean probability of .408 and an average two-period cumulation of .480. Looking at the graph in Figure B-4 we see that the best three beta function curve underestimates the twelve-issue cumulative audience by about ten percent when the curves are fitted with the values of the average audience and the two-period cumulation. The single beta function curve and other three beta function curves fall even further below the empirical cumulation curve. It is not clear why this is so. It may be that there are other problems with Schyberger's data than the problem

---

would generally use in a study like this and in fact, because of a secular trend in readership during the time of his study, his values are consistently too low. We have taken his values and averaged them in reporting them in the present study.

with the averaging in order to get the cumulation. It may also be that cumulation is simply quite different for data of Swedish magazines in Malmo and that the model does not fit. Another possibility, of course, is that the model will fit but that we simply have not been able to find the proper values of the parameters to generate the maximum cumulation given the two fitted points.

The final exploration in the effects of the various parameters was performed on hypothetical data which might be appropriate to newspaper cumulation values. Here the average audience for a hypothetical newspaper was taken as fifty percent of the population with a two-period cumulation of sixty-five percent of the population. The proportion of the population in the low distribution varied from ten to forty percent with a mean probability of exposure from .010 to .150 and two-period cumulation from .015 to .250. For the high distribution, the population proportion ranged from ten to thirty percent, with a mean probability of exposure varying from .800 to .900 and a proportional two-period cumulation from .900 to .970. For these very high values we have generated thirteen different cumulation curves and frequency curves, the extremes of which are shown in the graph of Figure B-5. The graphs show a significant variation in the extremes of the curves for the thirteen-period cumulation. The smallest value which we found exposed at least once was about eighty

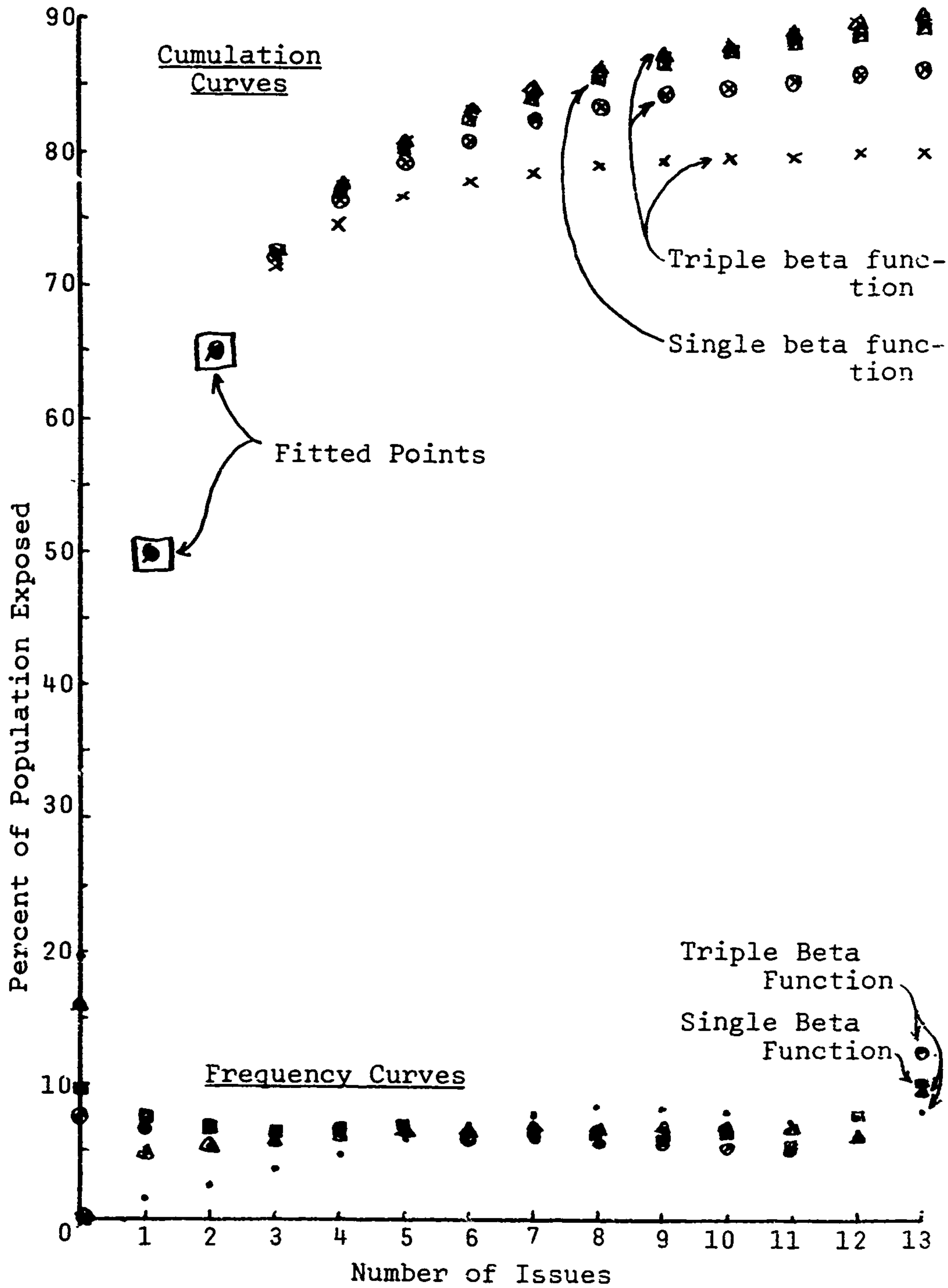


Figure B-5. Cumulation and Frequency Curves for Various Values of the Parameters for Paper

percent of the population, while the maximum exposure was approximately ninety percent of the population. Note that the range is about ten percent of the population or about twelve percent possible error in the thirteen-period cumulation.

If we look at the three cases together, we seem justified in making the following generalization: the greatest absolute and proportional errors occur, or are likely to occur, in those distributions for which the average audience and two-period cumulation are small. This is the case for the Better Homes and Gardens cumulation curve in which we see that the range of the cumulation curve can be quite large and seems to be steadily increasing for the one extreme curve as the time periods progress, at least through twelve time periods. On the other hand, for the data where the average audience and cumulation are larger, the errors or variations in the curves are smaller both absolutely and proportionately, and the rate of growth of error is smaller. In general, we might make a rough estimate that for the average newspaper vehicle our estimate of the cumulative number of people exposed may be in error by as much as ten percent after thirteen issues.[10] For a radio vehicle which

[10]There is, however, an asymptote to the cumulation curve, and therefore after many time periods, the cumulation error must diminish. This effect probably happens earlier for those vehicles with larger average audiences.

has a very small average audience, our estimate of the cumulative audience may be in error by as much as ten to fifteen percent, after twelve or thirteen issues. This analysis does seem to show however that for many vehicles the proper choice of parameters (were they known) would produce cumulation and frequency curves closely approximating the empirical curves. Only in the case of the Swedish magazines (discussed in more detail below) do we seem to have some regular deviation from the empirical curves.

The Fit of the One-and Three-Beta Function Models for Thirty-Two Vehicles

In order to explore more fully the feasibility of using the one-or three-beta function models, we have fitted the curves to each of thirty-two vehicles and in the table below we present for comparison the largest real cumulation value which has been measured for these vehicles and the comparable cumulation value derived from the one-and three-beta function models by fitting the average audience and the two-period cumulation.

In examining these figures we recall once more that the cumulation values generated by the one-beta function model are fixed once the average audience and the two-period cumulation are known. However, for the three-beta function model these two values do not fix the cumulation curve and it appears from our previous explorations that the values for twelve or thirteen time periods may vary by as much as ten or fifteen percent of the population (depending on the choice of the other parameters), or by one or two percent times the number of time periods beyond the second for which the values are calculated. Thus, the further the time period is beyond the second time period, the greater the variation between the values generated by the different parameters of the three-beta function model, at least up to thirteen time periods; a two percent error at time period thirteen may be considered a much smaller error than a two percent error at time period four. (For example, a two percent error at time period four may grow to a ten or twelve percent error at time period thirteen.) There are two other factors to keep in mind in examining the table. The first of these is that

Table B-1. Comparison of the Real and Derived Cumulations (Using the One and Three Beta-Function Models) for 32 Vehicles

Vehicle	Number of Time Periods	Empirical Cumulation	Comparable Values Derived from the One-Beta Function: Value	One-Beta Function: Error	or Three-Beta Function Models: Value	Three-Beta Function Models: Error
American Mass Magazines						
Gardens (1955)	12	35.7	34.2	-1.5	35.4	-0.3
This Week (1953)	6	35.3	36.3	1.0	37.8	2.5
Journal (1953)	6	23.9	23.2	-0.7	25.0	1.1
Journal (1958)H	6	40.4	40.0	-0.4	40.0	-0.4
Post (1953)	6	29.6	28.7	-0.9	30.6	1.0
Post (1958)H	6	42.1	40.3	-1.8	41.3	-0.8
Life (1950)	13	53.1	53.0	-0.1	57.3	4.2
Life (1953)	6	50.6	49.4	-1.2	52.0	1.4
Life (1958)	6	48.7	48.2	-0.5	49.4	0.7
Life (1958)H	6	60.5	59.1	-1.4	58.5	-2.0
Look (1953)	6	41.8	40.7	-1.1	43.2	1.4
Look (1958)H	6	52.7	51.4	-1.3	52.3	-0.4
Look (1961)	4	39.7	38.4	-1.3	38.5	-1.2
Digest (1958)H	6	62.2	60.8	-1.4	59.8	-1.4
American Baby Magazines						
American Baby (1964)	12	29.2	28.0	-1.2	30.7	1.5
Your New Baby (1964)	12	71.0	59.7	-11.3	65.0	-6.0
My Baby (1964)	12	60.5	51.9	-8.6	57.6	-2.9
Baby Talk (1964)	12	61.5	47.8	-13.7	48.3	-13.2

Table B-1. (Continued)

Vehicle	Number of Time Periods	Empirical Cumulation	Comparable Values Derived from the One-Beta Function: Value	One-Beta Function: Error	or Three-Beta Function Models: Value	Three-Beta Function Models: Error
American Television Programs						
Colgate (1953)	4	42.6	44.0	1.4	46.2	3.6
Texaco (1953)	4	38.0	38.9	.9	39.4	1.4
Skelton (1953)	4	39.2	40.0	.8	40.4	1.2
Show of Shows (1953)	4	42.3	43.6	1.3	43.7	1.4
Fireside (1953)	4	29.0	30.0	1.0	30.7	1.7
American Radio Programs						
Charlie McCarthy (1953)	4	25.0	25.1	0.1	27.3	2.3
Amos 'n' Andy (1953)	4	32.1	32.4	0.3	35.4	3.3
Jack Benny (1953)	4	35.1	34.9	-0.2	38.0	2.9
Lux Radio Theater (1953)	4	21.6	21.6	0.0	23.6	2.0
Swedish Magazines						
Allers (1962)	12	52.0	45.6	-6.4	45.3	-6.7
Hemmets Veckotidning (1962)	12	62.7	55.3	-7.4	53.8	-8.9
Se (1962)	12	71.5	63.9	-7.6	64.2	-7.3
Vecko-Revyn (1962)	12	58.3	51.4	-6.9	52.3	-6.0
Vi (1962)	12	43.3	41.9	-1.4	40.6	-2.7

the three-beta function model can always be made equal to the one-beta function model simply by giving each of the distributions the overall average audience and two-period cumulation. Thus, in the cases where the one-beta function model is better than a particular three-beta function model we must remember that the three-beta function model includes the possibility of generating exactly the same values as the one-beta function model. Finally, the values of the parameters in the three-beta function model were chosen on an intuitive basis by the researcher. These values are not necessarily the best choices and some experimentation would likely produce cumulation closer to the empirical calues, especially in the case of Swedish magazines.

The first group of vehicles are fourteen American mass magazines whose audiences were studied during the period from 1950 to 1961. For this group of magazines the cumulation was measured from four to thirteen issues. We see that the cumulations derived from either the one-beta function or the three-beta function model fit these data quite well. The errors in the one-beta function cumulation are generally quite small even for the twelve and thirteen time period cumulations with, however, one systematic error in that the model generally understates the cumulation by a very small amount. The three-beta function model with the parameters chosen in a rather ad hoc fashion by the researcher does reasonably well and if we recall the variation possible in the cumulation with a different choice of parameters it seems clear that the three-beta function model can be made to fit the empirical data

quite well. However, if the empirical data did not include the cumulations for the three distributions, it would seem that the best estimate of the parameters for the three-beta function model would be identical with the parameters for the one-beta function model.

The second group of vehicles are four American baby magazines. These are magazines which circulate primarily among the population of pregnant or recently pregnant American women and thus represent a somewhat esoteric type of vehicle. For these vehicles, the one-beta function model does not do well at all in modeling the empirical cumulation. The error seems to be much larger for those magazines which have a very large empirical cumulation. The three-beta function models do significantly better except in the case of Baby Talk for which the three-beta function model with the particular parameters chosen for this calculation does not do significantly better than the one-beta function model. We conclude from this group of vehicles, that the one-beta function model is not rich enough to model these magazines' audiences but that very likely the three-beta function model can come very close to modeling the audiences.

The next two groups of vehicles in the table are American radio and television programs whose cumulations were measured in the 1953 Politz Study of Four Media.[11] For these vehicles,

---

[11] Alfred Politz Research, Inc., A Study of Four Media: Their Accumulative and Repeat Audiences, (New York: Time Incorporated, 1953)

the cumulation has been measured through only four time periods, but for that cumulation both the one-beta function and three-beta function models seem sufficient. One must note, however, that the one-beta function and three-beta function models seem to systematically overestimate the cumulation and that this overestimation is consistently larger on the part of the three-beta function model. It would seem then that in modeling radio and television programs with no knowledge beyond the average audience and two-period cumulation, the researcher would do better using the three-beta function model as a one-beta function model than by picking ad hoc values of the parameters.

The final group of vehicles consists of five Swedish magazines which were studied by Schyberger in Malmo in 1962. For these magazines we have cumulation values through twelve issues. We have spoken previously about the need to adjust Schyberger's figures by averaging over many issues, a procedure which he did not use. This procedure has been used in calculating the average audience and two-period cumulation for these magazines. Nevertheless, the cumulation values derived from either the one-beta function or the three-beta function model did not seem to reproduce well the cumulations which were measured empirically; the derived values are consistently lower than the observed values and the larger the observed values, the more in error are the derived values. Even for Hemmets Veckotidning for which we explored a variety of parameters for the three-beta function model the fit is not good for any of the derived curves. Thus, for some reason the beta function models may not fit these data on Swedish magazines.

We conclude, then, that in the case of American mass magazines or American radio and television programs, the one and three-beta function models seem to fit the empirical data quite well with the three-beta function model offering a greater possibility of very closely fitting the empirical cumulation but, however, with the disadvantage that the eight parameters rather than the two parameters for one-beta function model must be calculated. For the case of American newspapers which, of course, are quite important in the context of this simulation we have no data on cumulation and can only point out that the three-beta function model allows a wide range of cumulation to be duplicated and that for vehicles with very large average audiences, as have the newspapers used in the simulation, the possibility of sizeable errors in cumulation is likely to be considerably lower.

The Prediction of the Two-Period Cumulation from the Average Audience

We have shown that in most cases we do reasonably well by fitting the parameters of the cumulation curve using the empirical values of the average audience and the two-period cumulation. We have also suggested (in Chapter V above) that it seems reasonable to expect some relationship between them. We have suggested above that this relationship might be of the following kind: that the larger the average audience, the smaller would be the proportional increase of the two-period cumulation over the average audience. This would be true not only because the base of the proportion, namely the average audience, is larger, but also because there remains a smaller proportion of the population to be newly exposed when t second issue arrives. In the graph below we have plotted the two-period relative accumulation (the ratio of the two-period cumulation to the average audience) against the average audience as a proportion of the population for thirty-two vehicles, including baby magazines, radio and television shows, Swedish magazines, and both household and individual audiences of American mass magazines. This graph shows clearly a very strong negative linear relationship between the relative cumulation and the proportion of the population in the average audience.

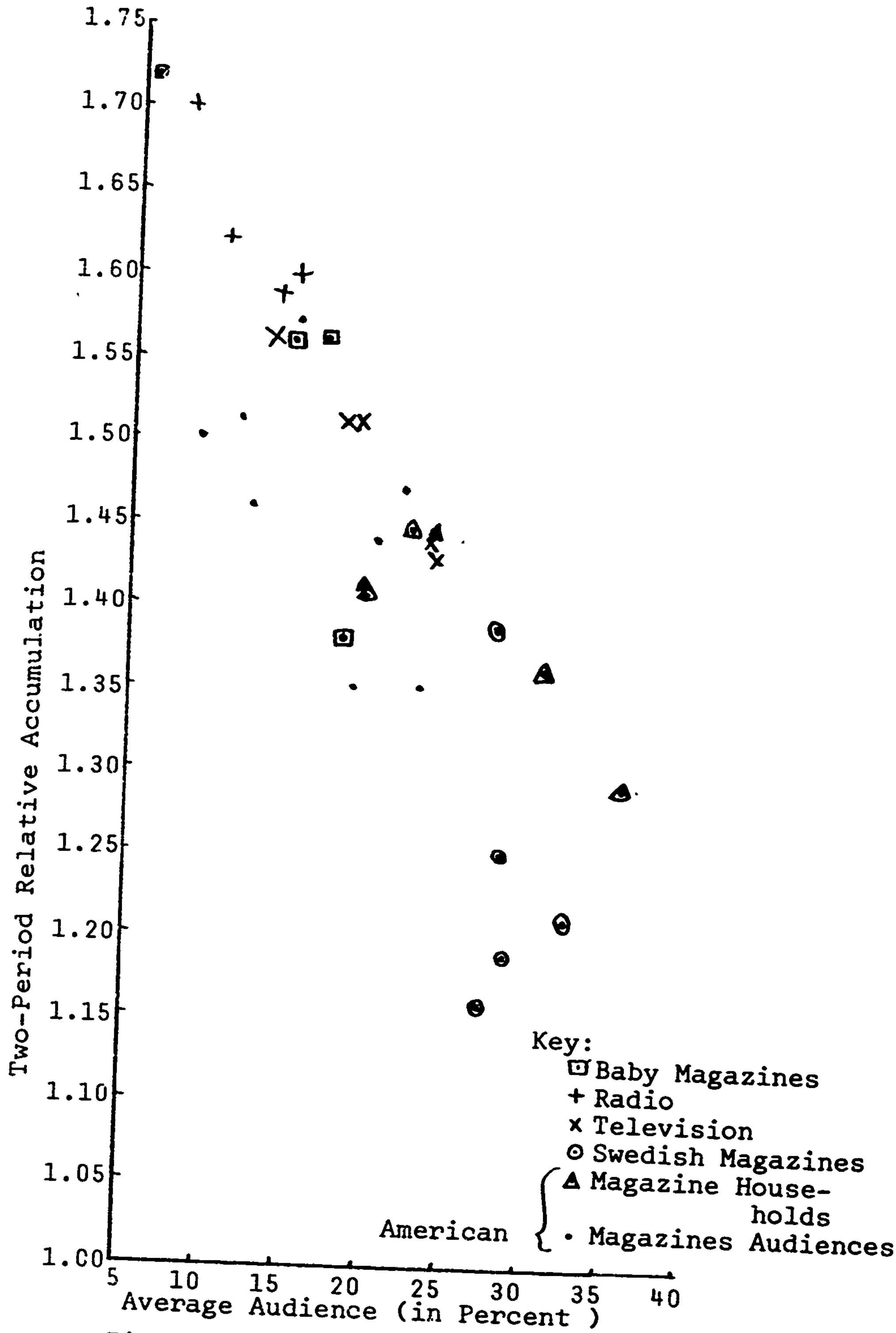


Figure B-6. The Two-Period Relative Accumulation As A Function Of The Proportion of The Population In The Average Audience, For Thirty-Two Vehicles

Although it seems logical to expect a relationship such as the one found in the graph of the thirty-two vehicles, we must at this point sound a warning. The two-period relative accumulation contains in both the numerator and the denominator the term C, which is the average audience Thus, even if the amount by which the two-period cumulation increases over the average audience were a random variable the fact that we add to this number the average audience and then divide the sum by the average audience and plot this quotient against the average audience would give us a high correlation and any statistics we might derive from this plot would be grossly misleading. Therefore, in running our correlations we have correlated the proportion of the population in the average audience with the proportional increase in the audience from the average audience to the two-period cumulation, Figure B-7 below shows a plot of this relationship. From this graph we see that the relationship is nearly linear up to an average audience of about twenty-two percent of the population, but that above this point the graph is not necessarily linear and in addition, the relationship becomes quite weak. In fact, the relationship may not hold at all for the cases of the Swedish magazines.

What form of relationship might we expect to find here? For the first graph where a linear relationship seems to apply, this relationship can be written in the general form of a linear equation as follows:

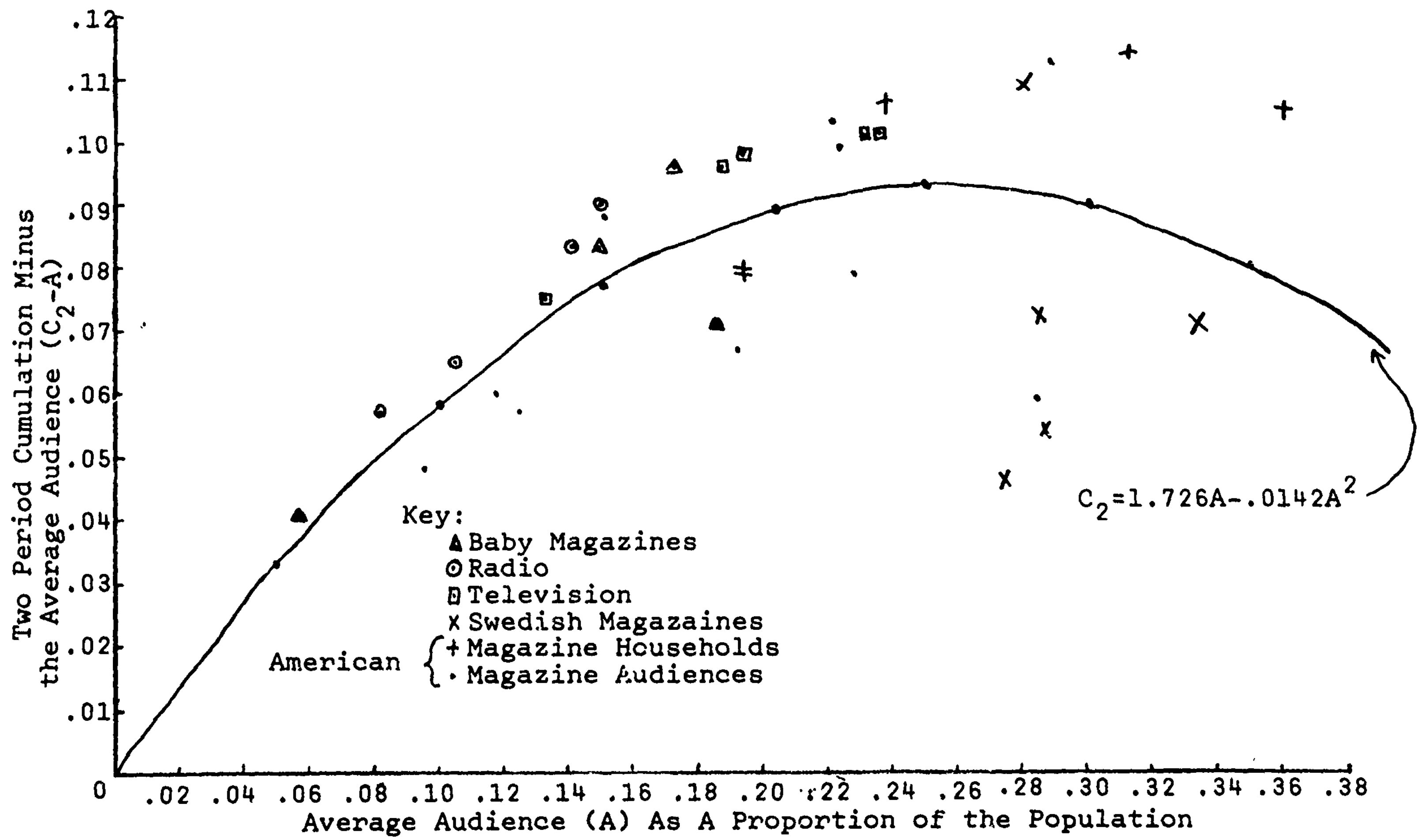


Figure B-7. The Difference Between the Two-Period and Average Audience Proportions as a Function of the Average Audience Proportion

Relative Accumulation = $C_1/C_2 = aC_1 + b$

where C_1 is the proportion of the population in the average audience, C_2 is the proportion in the two-period cumulation, and a and b are constants to be determined. If we then perform the operations on this equation which take the implicit value of the average audience from the left-hand side of the equation and leave as a dependent variable only the increase in the total audience from time period one to time period two as a function of the average audience, the equation then becomes a quadratic equation without a constant term. Thus, multiplying by C_1,

$$C_2 = aC_1^{\,2} + bC_1$$

and subtracting C_1 from both sides of the equation,

$$C_2 - C_1 = aC_1^{\,2} + (b-1)\ C_1$$

or

$$\text{net increase} = aC_1^{\,2} + b'C_1.$$

Therefore, we would make the prediction that if we were to run a regression using an equation of this type, we would find that the coefficients of the linear and quadratic terms were highly significant but that the constant term would not be significant. We have, in fact, run such a correlation and these findings happily are obtained. They are shown in Table B-2 below.

Table B-2. Regression Coefficients and Significance Levels

	Quadratic Term (C_1^2)	Linear Term (C_1)	Constant
Coefficient	$-.1117 \times 10^{-1}$	.5869	$.1424 \times 10^{1}$
Significance Level (one-tailed)	$0.025 > p > 0.010$	$.005 > p > .001$	$0.25 > p > 0.10$

The coefficients of both the linear and the quadratic terms are highly significant, but the constant term is not. Therefore, we conclude that we can predict the net increase from time period one to time period two in the proportion exposed from the equation

$$\text{Net increase} = 1.424 + .5869\, C_1 - .01117 C_1^2$$

with a multiple correlation coefficient of $R_{1.23} = .5851$ and a standard error estimate of 1.6429. If we wish to predict the proportional two-period cumulation directly from the equation we find

$$C_2 = 1.424 + 1.5869\, C_1 - .01117\, C_1^2.$$

The somewhat misleading multiple correlation coefficient is $R_{1.23} = .9817$[12] with a standard error of estimate of 1.6429.

---

12 Recall that this is an inflated value.

We can also perform the regression, forcing the constant term to be zero (recall that it was found to be not significantly different from zero) and find the resulting equation:

$$C_2 = 1.726\ C_1 - .0142\ C_1^2 \qquad \text{(B-21)}.$$

For this equation $R_{1.23} = .9814$ and the standard error of estimate is 1.6730.

Equation (B-21) is plotted in Fig. B-6. The negative coefficient of the squared term causes the required downturn from the peak of the curve at an average audience of about 26 percent. If the Swedish data points were deleted, this would probably occur at a somewhat higher value and the entire curve would probably be shifted slightly upward. At any rate, this curve does seem to offer a good prediction of the two-period cumulation from the average audience for American mass magazines, radio, and television programs for values of the average audience from about 8 percent to 25 percent of the population.

Addition of "Subscribers" to the Regression

We have mentioned one other logical variable for a regression predicting audience cumulation; namely, the proportion of the average audience that are subscribers or quite regular readers of the vehicle. The data for this variable were much more difficult to gather than the simple cumulation values and therefore, we have estimated it from either the proportion of copies going to subscribers, or from other data, for instance, the proportion of the audience exposed nine, ten, eleven, or twelve times out of twelve for a vehicle. For the twenty-two

cases where the proportion of the audience who are subscribers to the vehicle, or the proportion who are quite frequent users of the vehicle is known, we have introduced this variable into the multiple regression. We expect to find that the cumulation is a positive function of the proportion in the average audience, a negative function of the square of that proportion, and in addition, a negative function of the proportion of the average audience who are "subscribers" or heavy users of the vehicle. The results of the regression confirm this expectation. The individual coefficients are each highly significant and each bears the anticipated sign. The multiple correlation coefficient is $R_{1.234} = 0.7724$ ($R^2_{1.234} = 0.596$; $p=0.008$) and the unbiased estimate of the standard error of the values estimated from the regression is $S_{1.234} = 1.3313$. Thus, the addition of the estimate of the proportion of "subscribers" in the average audience adds significantly to our ability to predict the value of the two-period cumulation. (The inflated multiple correlation coefficient for this case is $R_{1.234} = .9875$).

An examination of the residuals shows that the absolute errors in the estimates for the Swedish magazine range from 0.77 to 2.02 standard errors, (only three of the other seventeen residuals are larger than .77 standard errors of estimate). It appears that the cumulation of audiences for these five magazines differs significantly from the American case <u>or</u> that the data measurement was not comparable. At any rate, the removal of these outliers from the data would probably somewhat increase the predictability of the American data.

We summarize the results of this section in the following points:

1. Every value of the relative accumulation which we have found (for electronic media and magazines in the United States and magazines in Sweden) lies within the interval 1.16 to 1.73.

2. We can make predictions of the proportion of the population in the two-period cumulative audience from the proportion in the average audience with a standard error of estimate of 1.6429 from the equation

$$C_2 = 1.4243 + 1.5896\,C_1 - .01117C_1^2.$$

3. We can make a better prediction using some estimate of the proportion of the average audience who are "subscribers", with a standard error of estimate of 1.3313, from

$$C_2 = -\,2.1211 + 2.1327\,C_1 - .02657\,C_1^2 - .02458\,(S/C_1).$$

4. These equations fit best for values of the average audience in the range from 5 to 25 percent of the population. Beyond this point, the functions do become monotonically decreasing as they logically should, but we have no data with which to check the fit.

5. The equations seem to hold well for electronic media and magazines in the United States. They do not seem to hold well for Swedish magazines and we have no data for newspapers anywhere. In addition, we must note that the data were gathered in the period from 1950 through 1962, and although we have no reason to believe that the world has changed in this respect since then, we must nevertheless be cautious about extrapolating these data to other time periods.

Table B-3. The Average Audience, Two-period Cumulation, Relative Cumulation, and Proportion of the Issue-copies Going to Individual Subscribers, for Fourteen Mass Audience Magazines

Magazine	Average Audience C_1	Two-period Cumulation C_2	Relative Cumulation C_2/C_1	Proportion of Issue-copies going to Individual Subscribers
American				
Life (1950)	20.3%[a]	29.2%[a]	1.438	65.2
Life (1953)	22.1	32.4	1.466	75.7
Life (1958)	22.3	32.2	1.444	
Saturday Evening Post (1953)	11.8	17.8	1.508	64.4
Look (1953)	15.1	23.9	1.583	70.3
Look (1961)	22.8	30.7	1.346	92.0
Ladies Home Journal (1953)	9.6	14.4	1.500	61.9
This Week (1953)	19.2	25.9	1.349	60.0
Better Homes and Gardens (1955)	12.5	18.2	1.456	64.9

Table B-3. (Continued)

Magazine	Average Audience C_1	Two-period Cumulation C_2	Relative Cumulation C_2/C_1	Proportion of Issue-copies going to Individual Subscribers
Swedish				
Allers (1962)	28.7	34.1	1.188	17.0
Hemmets Veckotidning (1962)	33.4	40.5	1.211	12.0
Se (1962)	28.0	38.9	1.389	14.0
Vecko-Revyn (1962)	28.5	35.7	1.252	15.0
Vi (1962)	27.5	32.1	1.164	100.0

[a]These columns represent percentages exposed of the total population.

Sources: The 1953 data for American magazines is taken from Alfred Politz Research, Inc., A Study of Four Media: Their Accumulative and Repeat Audiences (New York: Time, Incorporated, 1953); data for Better Homes and Gardens is taken from Alfred Politz Research, Inc., A Twelve-months' Study of Better Homes and Gardens Readers (New York: Meredith Publishing Company, 1955); for Look, from Audits and Surveys Company, Inc., Look Audience Study: 1961 (New York: Cowles Magazines and Broadcasting, Inc., 1961); for Life (1950) from Alfred Politz Research, Inc., A Study of the Accumulative Audience of Life (New York: Time, Incorporated, 1950); for Life (1958) from Alfred Politz Research, Inc., Life Study of Consumer Expenditures (New York: Time, Incorporated, 1958), 7 vols.; for Swedish magazines from Bo W:son Schyberger, Methods of Readership Research (Lund, Sweden: CWK Gleerup, 1965), pp. 93-117.

Table B-4. The Average Household Audience, Two-period Cumulation, and Relative Cumulation for Five Mass Audience Magazines

Magazine	Average Audience C_1	Two-period Cumulation C_2	Relative Cumulation C_2/C_1
Life (1958)	31.2%[a]	42.6%[a]	1.366
Saturday Evening Post (1958)	19.4	27.4	1.413
Look (1958)	23.8	34.4	1.445
Reader's Digest (1958)	36.0	46.5	1.292
Ladies' Home Journal (1958)	19.4	27.3	1.408

[a]These columns represent proportions exposed of the total population.

Source: Alfred Politz Research, Inc., *Life Study of Consumer Expenditures* (New York: Time, Incorporated, 1958), 7 vols.

Table B-5. The Average Audience, Two-Period Cumulation, Relative Cumulation, and Proportion of the Population Exposed to 3 or 4 of 4 Radio and Television Programs

Program	Average Audience	Two-Period Cumulation	Relative Cumulation	Proportion Exposed to 3 or 4 of 4 Radio and Television Programs
Radio				
"Jack Benny"	15.0%[a]	24.0%[a]	1.600	6.9%[a]
"Amos 'n' Andy"	14.1	22.4	1.589	6.9
"Charlie McCarthy"	10.5	17.0	1.619	4.7
"Lux Radio Theatre"	8.2	13.9	1.695	2.6
Television				
Colgate	23.6	33.7	1.428	16.1
Show of	23.1	33.2	1.437	17.5
Skelton	19.4	29.2	1.505	11.4
Texaco	18.7	28.3	1.513	11.6
Fireside	13.3	20.8	1.564	8.0

[a] The percentages in these columns represent proportions exposed of the total population.

Source: Alfred Politz Research, Inc. A Study of Four Media: Their Accumulative and Repeat Audiences (New York: Time, Incorporated, 1953)

Table B-6. The Average Audience and Relative Cumulation of Four Baby Magazines in the Population of Pregnant Females

Magazines	Average Audience C_1	Two-Period Cumulation C_2	Relative Cumulation C_2/C_1
"American Baby"	5.7%[a]	9.8%[a]	1.72
"Baby Talk"	18.6	25.7	1.38
"My Baby"	14.9	23.2	1.56
"Your New Baby"	17.2	26.8	1.56

[a] The percentages in this column represent proportions exposed of four baby magazines in the population of pregnant or recently pregnant females (n = 3,685,000).

Source: The data are taken from an Audits & Surveys, Inc., Target Market Study of Eight Baby Magazines, August, 1955.

APPENDIX C

INSTRUCTIONS FOR CONTENT ANALYSIS OF CINCINNATI NEWSPAPERS

We wish to locate the occurrences in the mass media of certain selected themes for the period September 15, 1947 through March 1, 1948. The particular themes are described below. Obviously a single news item may carry more than one theme, e.g. a story on the Palestine debate in the U.N. might emphasize both the peacekeeping role of the U.N. (Theme 1) and also the dissension and dispute among the Great Powers (Theme 2). In this case both themes would be recorded for this story.

Certain of the questionnaire items whose responses are especially likely to be affected by the presence of the themes in the mass media are reproduced below. The item numbers following each theme indicate the relevant questionnaire item or items.

<u>Themes relating to the U.N.</u>

1. <u>U.N. Peacekeeping.</u> Any message which <u>relates the U.N.</u> to peacekeeping, promoting harmony among nations, discouraging aggression, discussing world affairs peacefully, settling world problems, etc. Includes editorials

urging upon the U.N. some peacekeeping action.
Items 1a, 2c, 5, 8, 9, 10, 11, 12.

2. Dissension in the U.N. Messages discussing disagreements in the U.N. among the Great Powers (the U S., England, Russia), bickering, arguing, name-calling, lack of unity, too much talk and not enough action, etc. Note that the message must explicitly mention the U.N.
Items 1b, 2a, 8, 9, 10, 12a.

3. The Veto in the U.N. Security Council. Any message mentioning anything about the veto power in the U.N., threats to veto proposals, number of Russian vetos, etc.

If the story mentions that the veto can be used only in the Security Council, please check this on the code sheet.

If the message somehow implies that the veto power means that the big powers must agree on an action, please check this on the code sheet.
Items 7, 12d.

4. The U.N. and the Rights of Man. Any message from which the reader can infer that the U.N. is concerned about equal rights for all people everywhere, e.g. Eleanor Roosevelt's columns on the work of the U.N. Human Rights Commission.
Item 6a.

5. The U.N. and World Trade. Messages which cause the reader to conclude that the U.N. has a responsibility to further trade between nations. Reports of U.N. debates on trade, tariffs, etc.
Items 6c, 11b.

6. The U.N. and World Health Conditions. Messages which imply that one task of the U.N. is to improve health conditions in different parts of the world. Famine relief.

Note that UNESCO may be named in this message as the agency with this task. If the U.N. is not mentioned explicitly, please note this in the space provided.
Items 6b, 11a.

7. The Slogan of the U.N. Information Campaign. "Peace begins with the United Nations; the United Nations begins with you." Any message containing this slogan.
Items 16, 17.

8. Cincinnati Plan Sponsors. Any message concerning the American Association for the United Nations or the U.N. Association of Cincinnati and/or their efforts to promote understanding of the U.N.
Items 14, 15.

9. Explicit Explanation of the U.N. Any message concerned primarily with explaining the purposes and/or workings of the U.N.
Items 5, 6.

10. Satisfaction with the U.N. Any message conveying approval of the U.N. or satisfaction with its progress and policies. Comments generally favorable toward the U.N.

Items 8, 10, 11.

11. Dissatisfaction with the U.N. Messages which convey hostility toward or displeasure with the U.N. or dissatisfaction with its progress. Comments generally unfavorable toward the U.N.

Items 8, 10, 12.

Themes relating to other international issues

12. Control of the Atomic Bomb. Messages which relate to control of the bomb, maintaining the U.S. secrets, possible use of the bomb, likelihood of Russia's developing the bomb, etc. Messages which further the reader's concern and interest in the control of the bomb. Ominous warnings and predictions.

Items 1c, 2b, 6d.

13. Russian-American Relations. Any message reminding the reader of growing hostility between the U.S. and Russia, e.g. messages concerning Russian foreign policy moves whichare inimical to U.S. interests; or concerning Russian warnings and threats (as in the U.N. exchanges between Dulles and Gromyko).

Items 1a, 1b, 2a, 3, 4, 8

14. Threats to Peace. Messages which raise the spector of war; e.g. messages concerning Russian threats, present wars as in Palestine or Greece. Discussions of the problems of keeping peace and the likelihood of another war. Dire predictions.

Items 1a, 2a, 3, 4.

QUESTIONNAIRE ITEMS

1. "When you think of the problems facing the United States now, which ones come to your mind first? What other problems do you think of?"

International Problems	March, 1948	Sept., 1947
a) Another war, maintaining peace	46%	24%
b) Relations with Russia	29%	16%
c) Control of the atomic bomb	1%	4%
d) United Nations	2%	1%

2. "We'd like to know how much interest the public takes in a number of questions. For example, do you yourself take a keen interest, only a mild interest, or practically none at all in news about:"

	Proportions taking a "keen" interest	
	March, 1948	Sept., 1947
a) Our relations with Russia	68%	54%
b) The control of the atomic bomb	56%	51%
c) The United Nations	34%	31%

3. "Do you expect the United States to fight in another war within the next ten years?"

	March, 1948	Sept., 1947
Yes	73%	48%
No	15%	38%
Don't Know	12%	14%
	100%	100%

4. "Do you think we can count on Russia to meet us halfway in working out problems together?"

	March, 1948	Sept., 1947
Yes	10%	14%
No	80%	74%
Don't Know	10%	12%
	100%	100%

Changes in information about the United States

5. "What would you say is the main purpose of the United Nations organization?"

	March, 1948	Sept., 1947
Proportion unfamiliar with the United Nations:	28%	30%

(The following questions were asked only of those who are familiar with the United Nations.)

6. "As far as you know, is the job of the United Nations to . . ."

	Proportions answering each item correctly	
	March, 1948	Sept., 1947
a) See that all people everywhere get equal rights (yes)	60%	55%
b) Improve health conditions in different parts of the world (yes)	55%	50%
c) Increase trade between countries (yes)	50%	47%
d) Deal with disarmament and control of the atomic bomb (yes)	43%	46%
e) Set up a new world language to be used in all countries (no)	34%	38%
f) Work out peace treaties with Germany and Japan (no)	10%	10%

7. "Have you heard or read anything about the veto power in the United Nations?"

	March, 1948	Sept., 1947
a) Yes	37%	34%
b) Proportion who could explain the working of the veto in terms of big power unanimity	7%	7%
c) Proportion aware that the veto power could only be used in the Security Council and not in the General Assembly as well	8%	7%

Changes in opinion about the United Nations

8. "In general, are you satisfied or dissatisfied with the progress that the United Nations organization has made so far?

	March, 1948	Sept. 1947
Satisfied	29%	34%
Dissatisfied	33%	28%
Don't Know	10%	8%
	72%	70%

9. "Do you think the United Nations organization will succeed in spite of the disagreements that have come up among England, Russia and the United States, or do you think these disagreements are so serious that the United Nations organization will fail?"

	Proportions among entire sample		Proportions among those queried	
	March, 1948	Sept., 1947	March, 1948	Sept., 1947
Will Succeed	35%	43%	48%	62%
Will Fail	25%	17%	34%	24%
Don't Know	12%	10%	18%	14%
	72%	70%	100%	100%

10. "Some people say there are so many disagreements in the United Nations, that we would be better off to get together with other countries and work on international problems outside the U.N."

"Other people say that working through the U.N. is the best way to preserve peace."

"How do you feel about this? (Should we work separately with any countries that want to join us, or should we work mainly through the United Nations?)"

	Proportions among entire sample		Proportions among those queried	
	March, 1948	Sept., 1947	March, 1948	Sept., 1947
Work Separately	12%	10%	17%	14%
Through U.N.	55%	57%	77%	81%
Don't Know	5%	3%	6%	5%
	72%	70%	100%	100%

11. "What would you say are some of the good things the United Nations has done so far?"

	Proportions among entire sample		Proportions among those queried	
	March, 1948	Sept., 1947	March, 1948	Sept., 1947
a) Handling of overseas food and relief	9%	6%	12%	8%
b) Creates or is creating unity between nations, has ended isolationism, breaks down cultural or trade barriers	8%	7%	11%	10%
c) Handling of specific issues: Palestine, Greece, atomic bomb, UNESCO, occupation forces, etc.	6%	6%	8%	8%
d) Helps nations to get together to talk over problems; acts as a forum or sounding board	5%	5%	7%	8%
e) Has helped small nations, curbed big ones	3%	2%	4%	3%
f) Has or is trying to curb Russian power	1%	1%	1%	2%
g) Is creating better understanding between Russia and the rest of the world	*	*	*	*
h) Vague general approval of U.N.	8%	7%	10%	9%
i) U.N. has done nothing good so far	11%	9%	14%	12%
j) Don't know, unable to answer	29%	31%	41%	44%
	80%[1]	74%[1]	108%[1]	104%[1]

[1]These columns add to more than their respective totals of 72%, 70%, 100% and 100% because a few respondents gave more than one response.

The criticisms of the United Nations were:

12. "What would you say are some of the bad things about the United Nations so far?"

	Proportions among entire sample		Proportions among those queried	
	March, 1948	Sept., 1947	March, 1948	Sept., 1947
a) U.N. is failing, too much talk and not enough action, lack of unity, bickering, arguing	19%	13%	25%	18%
b) Russian power is not curbed enough; difficulties with Russia; Russia should be barred	11%	13%	14%	17%
c) Handling of specific issues: not enough relief to destitute countries, Palestine, Greece, failure to proceed with disarmament and control of atomic bomb	6%	3%	8%	4%
d) Big nations have too much power; should do away with veto; make all nations equal	3%	2%	4%	3%
e) U.S. doesn't have enough power; is being taken advantage of	1%	1%	2%	1%
f) Vague general disapproval of U.N.	5%	3%	7%	4%
g) Everything is bad about it; should never have joined, should leave it now	1%	1%	2%	2%
h) Nothing bad about U.N.; just growing pains	7%	5%	9%	6%
i) Don't know, unable to answer	27%	35%	37%	51%
	80%[1]	76%[1]	108%[1]	106%[1]

[1]These columns add to more than their respective totals of 72%, 70%, 100% and 100% because a few respondents gave more than one answer.

13. "Do you happen to belong to any groups or organizations or attend any meetings where they talk about world affairs like the United Nations? (If 'No') Is there any particular reason why you don't?"

	Proportions among entire sample		Proportions among those queried	
	March, 1948	Sept., 1947	March, 1948	Sept., 1947
a) Belong or participate	12%	11%	16%	15%
b) Haven't been asked to join, not familiar with any groups to join	3%	3%	4%	4%
c) Time: Haven't got the time, too busy, other activities interfere	13%	16%	19%	23%
d) Just not interested	2%	3%	4%	6%
e) Personal reasons: age, health, newness in community, etc.	9%	5%	11%	6%
f) Dislike belonging to any organization	1%	2%	2%	3%
g) Wouldn't do any good, unnecessary, none of my business	1%	1%	1%	1%
h) Opposed to United Nations	*	*	*	*
i) Don't know why; no particular reason	31%	29%	43%	42%
	72%	70%	100%	100%

14. "Do you happen to know the names of any groups or organizations here in Cincinnati that are trying to help the United Nations?"

	Proportions among entire sample		Proportions among those queried	
	March, 1948	Sept., 1947	March, 1948	Sept., 1947
Yes	10%	7%	13%	10%
No	62%	63%	87%	90%
	72%	70%	100%	100%

The groups named scattered over all possible kinds. About 2% cited either the American Association of the United Nations or the United Nations Association of Cincinnati.

15. "In your opinion, can these groups and organizations (groups and organizations that are trying to help the United Nations) do anything to help the United Nations in a practical way?"

	Proportions among entire sample		Proportions among those queried	
	March, 1948	Sept., 1947	March, 1948	Sept., 1947
Yes	46%	40%	65%	57%
No	12%	10%	16%	14%
Don't Know	14%	20%	19%	29%
	72%	70%	100%	100%

16. "During the last six months, have you:

	Proportions saying "Yes" among entire sample[1]	Proportions saying "Yes" among those queried[2]
a) . . . seen anything in the newspaper about the United Nations?	59%	83%
b) . . . heard any radio news programs about the United Nations?	53%	74%
c) . . . heard any short radio mentions of the United Nations between programs?	26%	36%
d) . . . seen any signs or posters about the United Nations?	21%	29%

---

[1]The complementary figures for those shown in this column are the differences between them and 72%. Thus 59% saw newspaper coverage of the United Nations, 13% did not, and 28% were not asked the question; and similarly for each item.

[2]The difference between each of these figures and 100% is the proportion not exposed to each of the media among those who were familiar with the United Nations.

	Proportions saying "Yes" among entire sample[1]	Proportions saying "Yes" among those queried[2]
16. Continued.		
e) . . . heard anything about the United Nations in church?	12%	17%
f) . . . read any leaflets or pamphlets on the United Nations?	10%	14%
g) . . . been to any meetings where the United Nations was talked about or discussed?	9%	12%
h) . . . seen or heard anything else about the United Nations?	21%	29%

The main sources of United Nations' coverage mentioned under the "anything else" category were nationally circulating periodicals, newsreels and personal conversations.

The exact items mentioned were:

	Proportions among entire sample	Proportions among those queried
a) Conversations, talking to people	7%	10%
b) Nationally circulating periodicals, magazines	7%	9%
c) Newsreels, movies	4%	5%
d) Through school children	2%	3%
e) St. Xavier University Pageant	1%	1%
f) Books	1%	1%
g) Full length radio programs other than news	1%	1%

17. "Have you ever seen or heard the slogan: 'Peace begins with the United Nations; the United Nations begins with you'? (If 'Yes") Do you recall where you saw or heard it?"

17. Continued

Yes		49%[1]
Radio	27%	
Posters or signs	5%	
Newspapers	4%	
Heard it used in conversations	2%	
Magazines or books	2%	
Heard it used at meetings	1%	
Newsreels	1%	
Pamphlets	1%	
Miscellaneous	2%	
Don't recall	10%	
No, don't recall		51%
		100%

Although a good many people remembered hearing this slogan, not all of them can be assumed to have understood it. We have no systematic evidence on this point, but we have on record one interview in which the respondent said, yes, she had heard it over and over again on the radio, but then went on to add, "I never did find out what it means."

[1]This figure is smaller than the total for the sources mentioned, because some respondents named more than one source.

CINCINNATI CODE SHEETS

NEWSPAPER ______________

CODER ______________

ARTICLE HEADER

DATE			U.N. THEMES														OTHER			PAGE						FORMAT							
ARTICLE NUMBER	MONTH	DAY	1. U.N. PEACEKEEPING	2 U.N. BICKERING	3. U.N. VETO	3. VETO IN SEC. COUNCIL	3. VETO-BIG POWERS	4. U.N. HUMAN RIGHTS	5. U.N. WORLD TRADE	6. U.N. WORLD HEALTH	6. UNESCO HEALTH	7. U.N. SLOGAN	8. CAMPAIGN SPONSORS	9. U.N. EXPLAINED	10. SATISFACTION W/ U.N.	11. DISSATISFACTION W/ U.N.	12. CONTROL OF A. BOMB	13. US-USSR RELATIONS	14. TREATIES/PEACE WARS	FRONT PAGE LEAD STORY	OTHER FRONT PAGE	2nd OR 3rd PAGE	EDITORIAL PAGE(S)	WOMEN'S PAGE(S)	OTHER INSIDE PAGES	PRINT	PICTURE	TOP OF PAGE	MIDDLE	BOTTOM	NUMBER COLS. WIDE	COL. INCHES LONG

SELECTED BIBLIOGRAPHY

Books and monographs

Abelson, Robert P. "Simulation of Social Behavior," in The Handbook of Social Psychology. 2nd edition. Edited by Gardner Lindzey and Elliot Aronson. Reading, Massachusetts: Addison-Wesley Publishing Co., 1968.

The Advertising Research Foundation. The Continuing Study of Newspaper Reading: 138-Study Summary. New York: Advertising Research Foundation, Inc., 1951.

Belville, Jr., H.M. The True Dimensions of the Radio and Television Audience. New York: National Broadcasting, Inc., 1949.

Bishop, Yvonne M.M. Multidimensional Contingency Tables: Cell Estimates. Ph.D. Thesis, Department of Statistics. Harvard University, Cambridge, Massachusetts. February, 1967.

Bunker, J.P., Forrest, Jr., W. H. Mosteller, Frederick and Vandam, L. D., eds. The National Halothane Study. A study of the possible association between halothane anesthesia and postoperative hepatic necrosis. Report of the Subcommittee on the National Halothane Study of the Committee on Anasthesia, National Academy of Sciences - National Research Council. Washington, D. C.

Buying Habits Survey: The Cincinnati Market, (1951). The Cincinnati Post. Cincinnati, Ohio: March, 1952.

Chappell, Matthew N., and Hooper, C.C. Radio Audience Measurement. New York: American Book-Stratford Press, Inc., 1944.

City Hooperatings: Cincinnati, Ohio, Fall-Winter Report: October, 1947--through February, 1948. New York: C. E Hooper, Inc., 1948.

The Flow of Retail Buying Traffic in Cincinnati 1947: Fifth Survey. Cincinnati, Ohio: The Cincinnati Post. March, 1952.

Hanna, Joseph F. Information-Theoretic Techniques for Evaluating Simulation Models. Michigan State University, 1969.(Mimeographed).

Kramer, John F. The Three Distributions Case: How to Choose the Proper Number of Probabilities from Each Distribution When There Are Three Distributions. COMCOM/Simulation Memo #33, January 27, 1966.

Lazarfeld, Paul A. and Kendall, Patricia L. Radio Listening in America: The People Look at Radio-Again. New York: Prentice-Hall, Inc., 1948.

Lazarfeld, Paul F., and Stanton, Frank N., eds. Communications Research 1948-1949. New York: Harper and Brothers, 1950.

Life Study of Consumer Expenditures. Alfred Politz Research, Inc. New York: Time, Incorporated, 1958.

Look Audience Study: 1961. Audits and Surveys Company Inc. New York: Cowles Magazines and Broadcasting, Inc., 1961.

Media Balance. Life Incorporated. 1965.

National Opinion Research Center. Report 37A of the National Opinion Research Center. Cincinnati Looks Again. Chicago: University of Chicago, 1948.

A National Study of Newspaper Reading. Audits and Surveys Company, Inc., Volume I.

A National Study of Newspaper Reading: The Functions of Newspapers for Their Readers. New York: Audits and Survey Co. Inc., 1961.

Nielson, Arthur C. New Facts About Radio Research. New York: A. C. Nielson Company, 1946.

Pool, Ithiel de Sola. "Use of Available Sample Surveys in Comparative Research," in Encyclopedia of Sociology. 1963.

Roos, Leslie. Intersurvey Comparison-A Progress Report. Massachusetts Institute of Technology, 1967. (Mimeographed.)

Sandage, C. H. Radio Advertising For Retailers. Cambridge: Harvard University Press, 1945.

Schramm, Wilbur. Mass Communications. 6th edition. Urbana: University of Illinois Press, 1949.

Schyberger, Bo W:son. Methods of Readership Research. Lund, Sweden: CWK Gleerup, 1965.

Seehafer, Eugene F., and Laemmar, Jack W. Successful Radio and Television Advertising. New York: McGraw-Hill, 1951.

Selesnick, Herbert J. Link: A Program that Groups Media Among Which There Exists a Non-Random Duplication. COMCOM/Simulation Memo #17, August 26, 1964.

Simulations Media Mix: Technical Description. New York: Simulmatics Corporation, 1962.

Some Information Measures for Testing Stochastic Models. Michigan State, 1967. (Mimeographed.)

Star, Shirley A., and Hughes, Helen MacGill. Report on the Cincinnati Plan.

A Study of Four Media: Their Accumulative and Repeat Audiences. Alfred Politz Research, Inc. New York: Time, Incorporated, 1953.

A Study of the Accumulative Audience of Life. Alfred Politz Research, Inc. New York: Time, Incorporated, 1950.

Target Market Study of Eight Baby Magazines. Audits and Surveys, Inc. August, 1955.

A Twelve Months' Study of Better Homes and Gardens Readers. Alfred Politz Research, Inc., Des Moines, Iowa: Meredith Publishing Company, 1956.

Journal articles

Abelson, Robert P. and Bernstein, Alex. "A Computer Simulation Model of Community Referendum Controversies." Public Opinion Quarterly, XXVII (1963), pp. 93-122.

Bailey, Gilbert. "To Make Us Aware." New York Times Magazine, March 7, 1948. pp. 24-5

Bauer, Raymond A., and Alice H. "America, Mass Society and Mass Media." The Journal of Social Issues, Volume XVI, No. 3, 1960, p. 12.

Birch, M.M. "Maximum Likelihood in Three-Way Contingency Tables." Journal of the Royal Statistical Society, Series B. Volume 25, 1963, pp. 220-33.

Brown, David T. "A Note on Approximations to Discrete Probability Distributions." Information and Control, Volume 2, 1959, p. 386.

Bush, Clinton R. "Content and 'Mise en Valeur': Attention as Effect." Journalism Quarterly, 1960, 37:3, pp. 435-6.

Deming, W. Edwards and Stephan, Frederick F. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known." Annals of Mathematical Statistics, Volume 11, 1940, pp 427-44.

Deming, W. E. and Stephan, F. F. "The Sampling Procedure of the 1940 Population Census." Journal of the American Statistical Association, Volume 35, 1940, pp. 615-30.

Hanna, Joseph F. "A New Approach to the Formulation and Testing of Learning Models." Synthese, Volume 16, 1966, pp. 344-380.

Herman, Charles F. "Validation Problems in Games and Simulations with Special References to Models of International Politics." Behavioral Science, Volume 12, 1967, pp. 216-231.

Hyett, G. P. Paper read to the Statistics Seminar, London School of Economics, February, 1958.

Howard, Ronald A. "Stochastic Process Models of Consumer Behavior." The Journal of Advertising Research, Volume 4, 1966, pp. 35-42.

Metheringham, Richard A. "Measuring the Net Cumulative Coverage of A Print Campaign." Journal of Advertising Research, Volume IV, No. 4 (December, 1964), pp 23-28.

Mosteller, Frederick. "Association and Estimation in Contingency Tables." Journal of the American Statistical Association, Volume 63, 1968, p. 19.

Sandage, Charles H. "Qualitative Analysis of Radio Listening in Two Central Illinois Counties." Bureau of Economic and Business Research Bulletin Series, No. 58 and University of Illinois Bulletin, Volume XLVI, No. 50 (March, 1949).

Schramm, Wilbur, and White, David M. "Age, Education, and Economic Status as Factors in Newspaper Reading." Journalism Quarterly, June, 1949, pp. 402-12.

Scott, William A. "Reliability of Content Analysis: A Case of Nominal Scale Coding." Public Opinion Quarterly, Fall 1955, pp. 321-25.

Star, Shirley A., and Hughes, Helen MacGill. "Report on an Educational Campaign: The Cincinnati Plan for the United Nations." American Journal of Sociology, January, 1950, p. 9.

"Students Study Local United Nations Programs." Cincinnati Times-Star, November 27, 1947, p. 45.

Swanson, Charles E. "What They Read in 130 Daily Newspapers." Journalism Quarterly, Fall, 1955, 32:4, pp. 411-421.

Tannenbaum, Percy H., and Lynch, Mervin D. "Sensationalism: The Concept and Its Measurement." Journalism Quarterly, 1960, 37:3, pp. 381-91.

"U.S. Video Sets Now 484,350." Variety, August 25, 1948.

BIOGRAPHICAL NOTE

The author attended Miami University at Oxford, Ohio, during the years 1955-57, 1958-59, majoring in physics, and received a B.A. in 1959. He attended the University of Illinois at Urbana during the years 1959-61, majoring in physics, and received an M.S. in 1961. He attended the Massachusetts Institute of Technology during the years 1961-69 and received a Ph.D. in Political Science in 1969.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Massachusetts Institute of Technology
Center for International Studies

2a. REPORT SECURITY CLASSIFICATION: unclassified

2b. GROUP

3. REPORT TITLE

A Computer Simulation of Audience Exposure in a Mass Media System: The United Nations Information Campaign in Cincinnati. 1947-48.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

John Francis Kramer

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1969	535	69

8a. CONTRACT OR GRANT NO. AF 49(638)1237 (ARPA)	9a. ORIGINATOR'S REPORT NUMBER(S) C/69-11
b. PROJECT NO. 9717	
c. 61101D	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) **AFOSR 69-2760TR**
d. 681313	

10. DISTRIBUTION STATEMENT

1. This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH, OTHER	**Air Force Office of Scientific Research 1400 Wilson Boulevard (SRLB) Arlington, Virginia 22209**

13. ABSTRACT

A simulation of Cincinnati mass media system predicts frequency and reach of flow of messages from known facts taken from census statistics, newspaper and radio audience studies, and a content analysis of the press relevant to attitudes and opinions measured by NORC survey of the effects of a public information campaign on the United Nations made in 1947-48. Three trial scenarios tested plausibility of exposure and consistency with which model synthesized input data before actual scenario of twelve real themes was run. Themes most closely related to large changes shown in NORC panel were those for which simulation predicted highest average frequency of exposure. Two models relate information or attitude change to frequency of exposure and correlations are sought to themes with changes in attitudes and information and recall of exposure in NORC panel. The distribution of exposure was relatively constant from theme to theme and correlations were small. Suggestions are made for increasing usefulness of model.

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Audience exposure						
Campaign						
Cincinnati						
Information						
Media						
NORC						
Simulation						

UNCLASSIFIED

Security Classification